

QY	1870	aLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVal	1890	Db	4119	CATGACGAGCAGCGGCACTCCCGAGTGTGTGTACCGCTGTGTATCCGGATGGGGA	4178
Db	3039	CCTCCAGCAGAGACCCAGTGTGTTTCAACAACACAGGCATGAAGGGTATCATCACCGCTGT	3098	QY	2250	uGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAl	2270
QY	1890	lGluProSerLysProThrValLeuArgSerThrSerSerProValArgProAl	1910	Db	4179	ACAGACGGAGCCAGCAGATGGGCTCCAAGTCTCCAGGCAACACACGACGCCGCCAGC	4238
Db	3099	GGAGCCAGCAGCCACGCTGCTGAGGTCCACTCCACTCTCCACCGTTCGCCACGC	3158	QY	2270	aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluI	2290
QY	1910	aAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPr	1930	Db	4239	CTTCTTCAGCAGCTGACCGAGAGCACTCCGCCATGTCAGTCCAAAGAACAAAGAT	4298
Db	3159	TGCCACATTCACCTGCCACCCACTGCCCTGGGGCGCACCTCGATGGGGTCTACCC	3218	QY	2290	eAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGl	2310
QY	1930	oThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluAr	1950	Db	4299	CAACAAGAAGCTGAACACCCACACCGGAATGAGCTGAATACATATATCAGCAGCCCTGG	4358
Db	3219	TACCTCATGGAGCCCGCTGTGCTGCCCAAGAGGCCCGCGGTGCGCGGCACAGCG	3278	QY	2310	yThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGl	2330
QY	1950	gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGl	1970	Db	4359	GACGGAGATCTTCAATATGCGCCCATCACCGGAACAGCGCTTATGACCTATAGAGCCA	4418
Db	3279	GCCCCGAGCAGACACCGCGCATGCTTCTCGCCNAGCCCCCAGCCGCTCCGGGTGGA	3338	QY	2330	nAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLe	2350
QY	1970	uProAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGl	1990	Db	4419	GGCGGTGACAGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGCACT	4478
Db	3339	GCCCGCTCTCCCCAGCAGAGGCTCGAGGCCCGGCCCTAGTGCCTCCTCTCTCTGG	3398	QY	2350	uMetGlyLysTyrAspGlnThrGluSerProProLeuSerSerProLeuAlaPheAsnPr	2370
QY	1990	yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAlaSerProAs	2010	Db	4479	CATGGGTAATATGACCACTGGGAAAGATCCCGCGCTCAGCGCCAATGCTTTTAAACCC	4538
Db	3399	CCAGCCACCATCGCCGACCCCTGCGAAGAACTCGACCTCACCCAGCCAGCCCGGA	3458	QY	2370	oLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSe	2390
QY	2010	pProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr	2030	Db	4539	TCTGAATGCCAGTGCAGCCTGCCCTGCTATGCCCATAACCGCTGCTGACGGACGGAG	4598
Db	3459	CCCCCGGGCCACCTGCCTCGGCTCGAGCCCGCCAGCCCGGAAAGACTCAAGTAAACC	3518	QY	2390	rAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSe	2410
QY	2030	oPheSerIleGlnLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPr	2050	Db	4599	TGACACACACACTCACCTCGCCAGTGGCGCGGGAAGGCCAAGGTCTCTGGCAGACCCAG	4658
Db	3519	CTTTTCCATCCAGAACTGGAACCTCCGTTCTCTGGGTATCCACGCGCAGCAGCTACACCC	3578	QY	2410	rSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerVa	2430
QY	2050	oGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLe	2070	Db	4659	CAGCCGAAAGAACCAAGTCCCGCGCCCGGCTTGGCATCTGGGGACCGGCCACCCCTGTGT	4718
Db	3579	CGAAGGGGTGAGCGCCGTCAGCCCTGTGAGCTACCCAGTCTGACCCACGACAGGGGCT	3638	QY	2430	lSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr	2450
QY	2070	uProLysHisLeuGluLeuLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGl	2090	Db	4719	CTCTCAGTGACATCGAGGGAGACTGCAACCGCCGACCGCTCACCAACCCCGCTGTG	4778
Db	3639	CCCCAAGCACCTGGAGAGCTCGACAAGAGCCACTGGAGGGGAGCTGCGGCCCAAGCA	3698	QY	2450	pGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProleuIleMetAr	2470
QY	2090	nProGlyProValLysLeuGlyGlyAlaAlaHisLeuProHisLeuArgProLeuPr	2110	Db	4779	GGAGGACAGCCCTCGTCCGAGGTTCACGCCATTTCCCATACACCCCTGATCATGCG	4838
Db	3699	GCCAGGCCCGGTGAAGCTTGGCGGGAGCGCCACCTCCACACCTGCGCGCCGCTGCC	3758	QY	2470	gLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyPr	2490
QY	2110	oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGl	2130	Db	4839	GCTGCAGCGGGGTGTCATGCTTCCCAACCCCAACCGGGCTCCCGGGGACGCGGCC	4898
Db	3759	TGAGAGCCAGCCCTGCTCAGCCCGCTGCTCCAGACCGCCCGAGGGGTCAAGGTCAACA	3818	QY	2490	oLeuAlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuLeuCysSerGlnTy	2510
QY	2130	nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi	2150	Db	4899	CCTCGTGGCCCCCACCACCGCTGGGACGAGGAGGCCCAAGCCACTGCTCTGCTCGCAGTA	4958
Db	3819	GCGGGTGTGTCACCTGGCCAGCACATCAGTGAGGTTCATCACAGGACTACACCCGCA	3878	QY	2510	rGluThrLeuSerAspSerGlu	2517
QY	2150	hHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSe	2170	Db	4959	CGAGACACTCTCCGACAGCGAG	4980
Db	3879	CCACCCACAGCAGCTCAGCGCACCCCTGCGCGCCCTCTACTCTCTTCCCTGGGGCCAG	3938	RESULT 7			
QY	2170	rCysProValLeuAspLeuArgProProSerAspLeuTyrLeuProProProAspHi	2190	S83390			
Db	3939	CTGCCCGCTCTGGACCTCGCGCGCCACCCAGTGACCTTACCTCCCGCCCGGACCA	3998	LOCUS			
QY	2190	sGlyAlaProAlaArgGlySerProHisSerGluGlyLysArgSerProGluProAs	2210	DEFINITION			
Db	3999	TGGTGCCCGCGCGCTGGCTCCCGCCACACGAGAGGGGCAAGAGGTCTCCAGAGCCAAA	4059	ACCSSION			
QY	2210	nLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGl	2230	VERSION			
Db	4059	CAAGAGCTCGGCTTGGGTGGTGGAGGACGGTATTGACCTGTGCTCCACCGGAGGG	4118	KEYWORDS			
QY	2230	yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGl	2250	SOURCE			
				ORGANISM			
				Hom sapiens (human)			
				Hom sapiens			
				Eukaryota; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;			
				Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
				REFERENCE			
				1 (bases 1 to 2930)			
				Authors			
				Sande, S. and Privalsky, M.L.			

Db 819 GGTCCAGCTCCAGTCCGCTACTCAGAGCATGCCAAGGCCCGCGTGGGCCCTGTCTACCAT 878
Qy 1150 tGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnG 1170
Db 879 GGGGCTGCCCTCCCTGATGACCCCAAAAGCTGGCACCTCTCAGCGGAGTGAAGCGGA 938
Qy 1170 uGlnLeuSerProArgGlyGlnAlaGlyProGluSerLeuGlyValProThrAlaG 1190
Db 939 GCAGCTGTCCCAACGGGGCAGGCTGGGCCACCGGAGAGCTGGGGGTGCCACAGCCCA 998
Qy 1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrly 1210
Db 999 GGAGGGCTCCGTGCTCAGAGGGGACAGCTCTGGGCTCAGTTCCGGGCGGAAGCATCACCA 1058
Qy 1210 sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh 1230
Db 1059 AGGCATTCCAGCACACGGGTGGCTTCGGACAGCGGCATCACATCCGGGCTCCATCAC 1118
Qy 1230 rHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyLys 1250
Db 1119 CCACGGCAGCGCCAGCTGACGTCTGTACAAAGGCGCACCATCACCGGCTCATCGCGAGGA 1178
Qy 1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy 1270
Db 1179 CAGCCCGAGTCCGTTCGACCGCGCGGAGGACAGCTGCCCAAGGGCCACGTCATCTA 1238
Qy 1270 rGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe 1290
Db 1239 CGAAGCAAGAGGGCCAGCTCTTGTCTATGAGGGTGGCATGCTGTGACCCAGTGTCTC 1298
Qy 1290 rLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgTh 1310
Db 1299 CAAGGAGGACGGCAGAAGCAGCTCAGGACCCCGCCATGAGCGCGCGCCCAAGCGCAC 1358
Qy 1310 rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db 1359 CTATGACATGATGAGGGCGCGCTGGGCGAGAGCATCTCTCTCAGCGAGCATCGAAGTCT 1418
Qy 1330 uMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisH 1350
Db 1419 CATGGCCGTGGCCATCCCGCGGAGGCGACACAGCCCGCCACCTCAAGAGCGACGACCA 1478
Qy 1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTy 1370
Db 1479 CATCCGCGGTCCATCACACAGGATCCCTCGTCTCTAGTGGAGGCGACAGGAGACTA 1538
Qy 1370 rLeuArgArgGluAlaLysLeuLysArgGluGlyThrProProProProProSe 1390
Db 1539 CCTGCGTCCGGAGGCCAAGCTCTAAAGCGGGAGGCGACGCTCCCGCCCGCCCGCCCTC 1598
Qy 1390 rArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
Db 1599 ACGGACCTTGACCGAGGCTCAAGACGCGAGGCGCTGGGCGCCCTGAAGCTGAAGCGGCG 1658
Qy 1410 aHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAr 1430
Db 1659 CCATGAGGCGCTGGTGGCCACGCTGAAGGAGGCGCGCTCCATCCATGAGATCCCGCG 1718
Qy 1430 sGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySe 1450
Db 1719 CGAGGAGCTGCGGCACACGCGCCGAGCTGCCCTGGCGCGCGCGCTCAAGGAGGGGCTC 1778
Qy 1450 rIleThrGlnGlyThrProLeuLysTyrAspThrGlyValAspThrThrGlySerLysLy 1470
Db 1779 CATACGAGGGCACCCCGCTCAAGTACGACCGGGCGGTCCACCTGGCTCCAAAAA 1838
Qy 1470 sHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLe 1490
Db 1839 GCACGAGCTAGCTCTCCCTCATCGCAGCCCGCGCGAGCTTCCACCCCGTGCACCCGCT 1898
Qy 1490 uAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSe 1510
Db 1899 GGATGTGATGGCGCGAGCCCGGCACTGGAAACGTGCTGTCTACGAGGAGAGCCTGAAGAG 1958

Qy 1510 rArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleVa 1530
Db 1959 CCGCCACGAGGACCGCCAGCAGCTCGGGGGCTCCATTGCGCGCGCGCCCGCTATTGT 2018
Qy 1530 lProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPh 1550
Db 2019 GCCTGAGCTGGGTAAGCCGCGCAGAGCCCCCTGACCTATGAGGACCAACGGGACCCCTT 2078
Qy 1550 eAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuG 1570
Db 2079 TCCCGGCCACCTCCACAGAGTTCCGCCGTGACCATCGGGAGCCACCGCGCGCTGCA 2138
Qy 1570 nGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProAr 1590
Db 2139 GGAGGGCAGCTTTCTGTCAGCAAGCATCCCGAGGACCGAAAGCTGACGCTGACGCGCTCG 2198
Qy 1590 gGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPr 1610
Db 2199 TGAGATCGCAAGTCCCGCACAGCACCGTGGCGGAGCACCCACACCCCATCTCGCC 2258
Qy 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLe 1630
Db 2259 CTATGAGCACCTGCTTCGGGGCGTGAGTGGCGTGACCTGTATCGGAGCCACATCCCCCT 2318
Qy 1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIleTyrTy 1650
Db 2319 GGCTTCGACCCACCTCCATACCCCGCGGCATCCCTCTGGAGCGCAGCGCTGCCTACTA 2378
Qy 1650 rLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIl 1670
Db 2379 CCTGCCCCGACACCTGGCCCCCAACCCCACTACCGGCACCTGTACCCACCTACCTCAT 2438
Qy 1670 eArgGlyTyrProAspThrAlaAlaLeuGluLeuArgGlnThrIleIleAsnAspTyrIl 1690
Db 2439 CCGCGCTACCCCGCACACGGCGCGCTGGAGAACCGGACAGCATCATCAATGACTACTAT 2498
Qy 1690 eThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLe 1710
Db 2499 CACCTCGCAGCAGATGCACACACACGCGCCACCGCCATGGCCCGCAGCTGATATGCT 2558
Qy 1710 uArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGl 1730
Db 2559 GAGGGGCTCTCGCCCCGCGAGTCTCTCGTGGCACTCAACTAGCTGCGGGTCCCGGAG 2618
Qy 1730 yIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyTh 1750
Db 2619 CATCATCGACCTGTCCCAAGTGCACACCTGCTGTGCTCGTGCCTCCCGCCGACACGAGC 2678
Qy 1750 rProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSe 1770
Db 2679 CCCAGCCACCGCATGGACCGCTTGCTTACCTCCCGCCACCGCGCCCGCTTCAGCAG 2738
Qy 1770 rArgHisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrTh 1790
Db 2739 CCGCCACAGCAGCTCCCACTCTCCCGAGGAGTCCACACACTTTCACAAAACCAACCCAC 2798
Qy 1790 rThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArgGluAr 1810
Db 2799 CAGCTCTCGTCCGAGCGGAGCGAGACCGGATCGAGCGGAGCCCGGATCGGAGCG 2858
Qy 1810 gGluLysSerIleLeuThrSerThrThrValGluHisAlaProIleIleTrpArgProGl 1830
Db 2859 GGAAAGTCCATCTCCATCCCGTCCCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 2918
Qy 1830 yThrGluGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1850
Db 2919 TACAGAGCAGAGCAGCGGAGCAGCGAGCGCGCGGGGTGGGGGAGCGAGCGAGCGCG 2978
Qy 1850 gProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAl 1870
Db 2979 CCGCGCTCCCACTCCCATGCGCCACGAGCAGCTCGCCCATCTCCCTTCGGAGCCAGGATGC 3038

```
Db 6857 TGTCCTCAGTACACTCAGAGGGGAGTCAATGCGGACACCACTCACCAACCGTGT 6916
Qy 2449 lTtpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMe 2469
Db 6917 GTGGAGGACCGGCGCTCATCTCAGAGGTCCAGCCATTCCCTACCAACCCCTTGATTAT 6976
Qy 2469 tArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySerGl 2489
Db 6977 GAGGCTACAGGAGGTGTCTATGCGCTCCCGCCGCCACCTGGCTTCGGCAGGAGCGG 7036
Qy 2489 yProLeuAlaGlyProHisHisAlaTtpAspGluGluProLyseProLeuLeuCySerGl 2509
Db 7037 GCCCTAGCTGTGCTCCACACAGCCTGGGATGAGGAGGCCCAAGCCACTGCTGTGTTTACA 7096
Qy 2509 nTyrGluThrLeuSerAspSerGlu 2517
Db 7097 GTATGAGACACTCTCGGACAGCGAG 7121

RESULT 6
HSU371146 5989 bp mRNA linear PRI 31-OCT-1995
LOCUS Human silencing mediator of retinoid and thyroid hormone action
DEFINITION (SWRT) mRNA, complete cds.
ACCESSION U371146
VERSION U37146.1 GI:1045654
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5989)
AUTHORS Chen, J.D. and Evans, R.M.
TITLE A transcriptional co-repressor that interacts with nuclear hormone
receptors
JOURNAL Nature 377 (6548), 454-457 (1995)
MEDLINE 96008552
PubMed 7566127
REFERENCE 2 (bases 1 to 5989)
AUTHORS Chen, J.D. and Evans, R.M.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-1995) J. Don Chen, Gene Expression Lab, The Salk
Institute, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
1..5989
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HeLa"
1..5989
/gene="SMRT"
496..4983
/gene="SMRT"
/notes="transcriptional co-repressor"
/codon_start=1
/product="silencing mediator of retinoid and thyroid
hormone action"
/protein_id="AAC50236.1"
/db_xref="GI:1045654"
translation="MEADADHPDKFAFAEAQKLPDPPCWTSGLPFPVPPREVIKAS
PHADPFAFYPFPHLPGLHDTARVLPFRPTISNPPPLISSAKHPSVLEKQIGA
ISQGSVOLHYVSEHAKAPVGVMTGLPMDPKLAPFSGVKQEQLSRPGQAPPE
SLGVPTAQASVLRGTALAGSVPGGISTKGIPSTRVPSDSAITVRSITHTGTPADLYK
GTITRIIGEDSPRLDGRDLSLPKHGVIYEGKGHVLSYEGGMSVTCSEKEDGRSS
GPETHAARKTYDMGRCVRRAISSIEGLMGRAPRPHSHPHLKEQHILRGSI
QGIPIRVYAEQEDYLREAKLKEGTPPPPPSRDLTEAYKTALQGLKLPKAEGL
VATVKEGRSITHEPRELRTPELPAPRLEKGSITQGPLKIDTGAISTTGKGD
VRLSGSGRTFPFVPHLDVNADARALACYEESLKSPTGATSSSGSSTARAPVIV
PELGRPROSLPTDHDGAPFAGHLPGRSPVTMRPTPLQELSSSSKASQDRKLST
PREAKSPHSVPBHHPHIPISPYEHLLRGVGLYRSHIPLAFDPTSIIRGIPLDA
AAYLPHRLANPNYPHLYPVLIRGYDPTAALENROTIINDYITSOQMHNTATMA
QYDMLRGLSPRESLALNAAAGPRGIIDLSQVPHLPVLVPTPTGTATAMDLYLP
TAQPFSSRHSPLSPGPGPHLTKPTTSSSRDRDRDRDREREKSLITSTTT
VEHAPIWRPGTEQSGSGSGSGSSSRPASHAHQHSPISPRTDQDALQORPSVL
```

exon

4483..4620

/gene="SMRT"

/note="alternatively spliced insert"

ORIGIN

Alignment Scores:

Pred. No.:	9,78e-152	Length:	5989
Score:	7885.00	Matches:	1510
Percent Similarity:	94.34%	Conservative:	7
Best Local Similarity:	93.91%	Mismatches:	14
Query Match:	59.67%	Indels:	77
DB:	9	Gaps:	7

US-09-522-753-5 (1-2517) x HSU371146 (1-5989)

Qy	984	ProProGluAspAlaAlaProThrLysProAlaProProAla	-----	998
Db	168	CCACCAAGA-----TCACCATATGAGCACCAGCCCCCTGCATAGGAATAATCATAT	218	
Qy	999	ProProProPro	-----	1002
Db	219	CCGAGCCACCTGTGGTTTCATAAGCCATCTGCCATTCCTGATGGGCTCCATGTATCTATT	278	
Qy	1003	-----GlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGly	1016	
Db	279	GCAGAGTCCCAAGTTTCATCAGCACTGAAACCAACCATCCGTCGTCGGAGAGGGAGGTGT	338	
Qy	1017	-----SerSerPro	-----	1020
Db	339	CTCCCTCTCTGTGCATAGGCATGAGGTCTCCCTCTAGTGGTGGTGGTGGAGGAGA	398	
Qy	1021	GlyLysSerArgSer	-----	1030
Db	399	GGAAGATTCCGAGCTCTGCTACCAACCCCGCGCTCGTCGGAGAGGGAGGTGT	458	
Qy	1030	-----	-----	1030
Db	459	CTTCGACGAGGGCTCATATCATATATCTTCTAGATGGAGGATGGAGCCGCCACCC	518	
Qy	1031	-AspLysGluAlaAlaAlaGluAlaGlnLysLeuProGlyAspProCysTrpTh	1050	
Db	519	CGAAGAGAGGCTTCGAGCGAGGCCAGAGCTGCTGGGAGCCCTTGTCTGGAC	578	
Qy	1050	rSerGlyLeuProPheProValProArgGluValLileLysAlaSerProHisAlaPr	1070	
Db	579	TTCCGGCTCTCCCTTCCCGTGCCTCCCTGAGGTGATCAGAGCTCCCGCATGCC	638	
Qy	1070	oAspProSerAlaPheSerTyrAlaProGlyHisProLeuProLeuGlyLeuHisAs	1090	
Db	639	GGACCCCTCAGCTCTCTACCTCCCTGAGTCCCTGAGTCCCTCCCGCATGCC	698	
Qy	1090	pThrAlaArgProValLeuProArgProProThrLileSerAsnProProLeuLileSe	1110	
Db	699	CATGCGCGCGCGCTCTCGCGCGCCACCCACCATCTCCACCCCGCTCCCTCATCTC	758	
Qy	1110	rSerAlaLysHisProSerValLeuGluArgGlnLileGlyAlaLileSerGlnGlyMetSe	1130	
Db	759	CTCTGCCAAGACCCAGCGCTCTCGAGAGGCAATAGGTGCCATCTCCCAAGAAATGTC	818	
Qy	1130	rValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMe	1150	

QY	1714	rProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLe	1734	Db	5780	CTCCAAACCTCTGGAAAGAGCTAGAGAAGGCCACTTGGAAAGGGAGCTGGGCACAAGCA	5839
Db	4760	ACCGCAGAGTCTCTCGCTGGCCCTCAATATATGCGCTGGCCCAAGAGGCATTATCGACCT	4819	QY	2090	nProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuPr	2110
QY	1734	uSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAl	1754	Db	5840	GCAGGCCCCATGAAGCTCAGCGGAGGCTGCCCATCTCCACCATCTCGGCGCACTGCC	5899
Db	4820	GTCCCAAGTGGCCACACTCGCGCTGTGTGTGCACCAAGCCAGGCACCCCTGCCACCGC	4879	QY	2110	oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisG1	2130
QY	1754	aMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSe	1774	Db	5900	CGAGAGCCAGCCCTCATCCAGCCCATCTCTCAGACTGCCCGAGGCATCAAGGTCAACA	5959
Db	4880	CATCGACCGCTTCCTACTCTCCCACTGCGCCCCCACCCTTCAGCAGCGCCACAGTAG	4939	QY	2130	nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi	2150
QY	1774	rSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSe	1794	Db	5960	GAGGCTGGTGCACCTCGGCTCAGCATCAGCAGAGGTCAATTACGCAGACTACACGCGCCA	6019
Db	4940	CTCACGCTGTCTCCAGAGGCCCACTCACTAGCTAAACCAACTGCCACATCTTCATC	4999	QY	2150	sHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSe	2170
QY	1794	rGluArgGluAArgAspArgGluArgAspArgAspArgGluArgGluLysSerIl	1814	Db	6020	CAACCCGAGCAGCTCAGTGGCCCCCTTCCCGCCCTCTCTACTCTTCTCCCGGAGCCAG	6079
Db	5000	GGAGCGGGAACGGGAACGTGAGCGGGAACGAGAC-----AAGTCCAT	5041	QY	2170	rCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHi	2190
QY	1814	eLeuThrSerThrThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSe	1834	Db	6080	CTGCCCTGTCTGGATCTTCGCGGCCACCCAGTGACCTTACCTCCCAACCCCGCAGCCA	6139
Db	5042	CCTCAGCTTACCATCTACAGTGAGCATGCACCCATCTGAGACCTGTGTACGGAGCAGAG	5101	QY	2190	sGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAs	2210
QY	1834	rSerGlySerSerGlySerSerGlyGlyGlyGlySerSerArgProAlaSerHi	1854	Db	6140	TGGCACCCCGAGCCCGGGATCCCCACAGTAGTGAAGGGGGCAAAAGGTCCCCAGAACCCAG	6199
Db	5102	CAGCGGGCT-----GGGGGCAGCAGCGCGCCCGCTCCCA	5137	QY	2210	nLysThrSerValLeuGlyGlyGlyAspGlyIleGluProValSerProProGluG1	2230
QY	1854	sSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnInAr	1874	Db	6200	CAAAATCTCGTCTCTGGGAGCAGCAGAGATGCCATTGAGCCTGTGTCTCCACAGAGGG	6259
Db	5138	CACC-----CACCAGCACTCGCCCATCTCCCCCGGACCCAGAGCGCTTGCAGCAGAG	5191	QY	2230	yMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyG1	2250
QY	1874	gProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLy	1894	Db	6260	CATGACTGAGCCAGCAGCATGCTCGGAGCACTGCGTACCCTGCTGTATCGAGACGGGA	6319
Db	5192	GCCCATGTGTGTGCACACACGAGCATGAAGGGCGTGGTCACTCCGTGGAAACCCCGCAC	5251	QY	2250	uGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAl	2270
QY	1894	sProThrValLeu-----ArgSerThrSerThrSerSerProValArgProAlaAl	1911	Db	6320	ACAGGCGAGAGCCC---AGGATGGGTCTAGAGTCTCCAGGCAACACAGCAGCCGCGCAAC	6376
Db	5252	GCCACCGTCTGAGGTGGGCCAGGTCCACCTCCACTCTTCGCTGTCCGCGCAGCTGC	5311	QY	2270	aPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGlu1	2290
QY	1911	aThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProTh	1931	Db	6377	CTTCTTCAGTAGCTGACTGAGAGCACTCCCGCATGTGTGAAGTGCAGAGCAGAGAT	6436
Db	5312	CACATTCCCACTGCCACCCACTGCCCACTTGTGTGGCACCTTGAAGGGGTCTTACCCTAC	5371	QY	2290	eAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProG1	2310
QY	1931	rLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPr	1951	Db	6437	CAACAAGAAACTCAACACCCACAACCGGAACGAGCCAGATACAATATTGGCCAGCCTGG	6496
Db	5372	CCTCATGGAGCCGCTCTGTATCCCAAGGAGACTCTCGGCTCGCCGCGCCGAGCGGC	5431	QY	2310	yThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerG1	2330
QY	1951	oArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPr	1971	Db	6497	GACGGAAATCTTCAACATGCGCGCATCTCTGAGCAGCCCTTATGACCTGTAGAAGCCA	6556
Db	5432	CGGGTGGAGCGCTGGCCATGCCCTTCTTACCAAAACCCCGGGCGG-----GAGCC	5482	QY	2330	nAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLe	2350
QY	1971	oAlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHi	1991	Db	6557	GGCGGTGCAAGAAACACCGCCAGCACCACATGGGGCTAGAGGCCATTATTAGAAAGGCACCT	6616
Db	5483	CGCCTCTTCAACAGCAAGAGCTCCGAGCCCGCATCCCTTAGCACCCCGCCAGCTCCAGCCA	5542	QY	2350	uMetGlyLysTyrAspGlnTrpGluSerProProLeuSerAlaAsnAlaPheAsnPr	2370
QY	1991	sAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspPr	2011	Db	6617	CATGGTAAATATGATCAGTGGGAAGAGCCCGCGCTCGGCGCCAATGCTTTTAACCC	6676
Db	5543	CACAGCCATCGCCCGCACCCACCAAGAACCTTGCACCCACCATGCCAGTCCGAGCCC	5602	QY	2370	oLeuAsnAlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyAr	2389
QY	2011	oProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPh	2031	Db	6677	TCTGAATGCCAGCGCCAGTCTGCCGCTGTGTCATGCCCATTAACCACTGCTGACCGAGC	6736
Db	5603	GCCGGCG---CCCACTCGGCCCTCAGATCTGCACCGCAGAAAGACTCAAAAGTAAACCTT	5659	QY	2389	gSerAspHisThrLeuThrSerProGlyGlyLysAlaLysValSerGlyArgPr	2409
QY	2031	eSerIleGlnGluLeuLeuArgSerLeuGlyTyrHis---GlySerSerTyrSerPr	2050	Db	6737	GAGTGACACGCACTCACTCTGCCAGTGGAGGTGGGAAAGCCCAAGGTCTCTGCGAGACC	6796
Db	5660	TTCATTCAGGAATTGAATCTCGTCTCTGGGTATACCAGTGGAGCTGGCTACAGCCC	5719	QY	2409	oSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSe	2429
QY	2050	oGluGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLe	2070	Db	6797	TAGCACCCGAAAGCCCAAGTCGCGCAGCACAGGCTTAGCGTCCGAGAGCCGACCCCTTC	6856
Db	5720	CGATGGGTGGAGCCCATCAGCCCGGTGAGCTCCCCCAGGCTGACCCACGACAGAGGGCT	5779	QY	2429	rValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgVa	2449
QY	2070	uProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysG1	2090				

Db 2660 GGACAGTACCCCAAGCCAGTTCCTCCCTGTGCTCCACCCAGCAGCACCTACAGCC 2719
Qy 1007 oGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAl 1027
Db 2720 AGAGGGTACGTCTCTCAGCAGTCGGGAGGAGTCCACGTGGCAAGTCCTCCGAGCCAGT 2779
Qy 1027 aProProAlaAspLysGlu-----AlaPheAlaAlaGluAl 1039
Db 2780 GCCTCTCCCGAGAAAGCAGCAGAAACCCCGCATCTCTTCCTCCGCTTTCCTCAACTGAGGG 2839
Qy 1039 aGlnLysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProValProPr 1059
Db 2840 CCCAAGCTACCGACTAGAGCCCCCAGCCAGCTGTCTATCGGGCTTCCCTTCCCATCTCCCTC 2899
Qy 1059 oArgGluValLysAlaSerProHisAlaProAspProSerAlaPheSerTrpAlaPr 1079
Db 2900 ACGGAGGTATCAAGACTTCCCCACACGCGCGCTGACCCCTCTGCTCTCTCTCTACACAC 2959
Qy 1079 oProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPr 1099
Db 2960 CCCCGGTACCCGCTGCTCTGGGCTTCCACGATAGTGGCCCGCGCTCTGTCACAGTCC 3019
Qy 1099 oProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGl 1119
Db 3020 CCCC--ATCTCTAAACCCCCACCCCTCATCTCTCTGCTCCCAAGCATCCCGCGTACTTGA 3076
Qy 1119 uArgGlnIleGlyAlaIleSer---GlnGlyMetSerValGlnLeuHisValProTyrSe 1138
Db 3077 GAGCAGCTGGTGGTCCCATCTCCACAGCAGGGGATGTCACTCCAGTCTGCTGTGCTCACTC 3136
Qy 1138 rGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspPr 1158
Db 3137 AGCATGCTCCAGGCCCTCCAGGCTCTCTCACATGGGCTGCCCTTCCCGTGGACCC 3196
Qy 1158 oLysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAl 1178
Db 3197 TAAGAGCTG----- 3206
Qy 1178 aGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyTh 1198
Db 3207 -----GGGAC 3211
Qy 1198 rAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPr 1218
Db 3212 AGCATGGCTCCCGCCACCATGGAAGCATCACCAAGGCTCCCCAGTACCCGGGTGC 3271
Qy 1218 oSerAspSerAlaIleThrThrArgGlySerIleThrHisGlyThrProAlaAspValle 1238
Db 3272 AGACGGCCCC-----AGCTACAGAGGCTCTATCACCCACGGCACGCGCGCAGACGTCT 3325
Qy 1238 uTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGl 1258
Db 3326 CTACAAGGGTACCATCAGCAGATCGTGGTGAGACAGCCCAAGTGCCTTGACCCGGC 3385
Qy 1258 YArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysGlyHisValle 1278
Db 3386 ACAGAGACACCTGCCCCAAGGGCCATGTCTATGAGGGCAAGAAAGCCACGTCT 3445
Qy 1278 uSerTyrGluGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSe 1298
Db 3446 ATCTATGAAGGTGTATGTCGTCTACAGTGTCTTAAGGAGGATGGAAGGACAGCTC 3505
Qy 1298 rGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgVa 1318
Db 3506 GGGCCACCCCATGATGATGCGGCCCTTAACGACCATATGATGATGAGGGCCGTGT 3565
Qy 1318 lGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGl 1338
Db 3566 AGCAGGACTGTCACTCAGCCAGCATAGAGGACTCATGGCCCGCGCATC---CCTGA 3622
Qy 1338 uArgHisSerProHisHisLeuLysGluGlnHisIleArgGlySerIleThrGlnGl 1358
Db 3623 GCAGCACAGCCCC---CACCTCAAGGAGCAGCATCACATCCGAGGCTCCATCAGCAGAG 3679

Qy 1358 yIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLe 1378
Db 3680 CATCCCGAGTCTCTATGTGGAGCGCAGAGGACTACTTACGGCGGAGGCCCAAGCTCTT 3739
Qy 1378 uLysArgGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyrIly 1398
Db 3740 GAAGCGAGAGGACACACACACCCACCCACCTCGGACCTGACTGAGACTACAA 3799
Qy 1398 sThrGln-----AlaLeuGlyProLeuLysLysProAlaHisGluGlyLeuVa 1415
Db 3800 GCCCGGCGCCCTGGACCTCTGGTCCCTGAAGCTGAAGCCGACTCAGAGGGTGTGT 3859
Qy 1415 lAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHi 1435
Db 3860 AGCAACTGTGAAGAGGAGGGCGGCGCTCTATCCATGAGATCCCGAGAGGAGCTGCGCG 3919
Qy 1435 sThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyTh 1455
Db 3920 CACACTGAGCTACCCCTGGCACACCGGCTCTGAAGGAGGGTTCATCACCACGGGCAC 3979
Qy 1455 rProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSe 1475
Db 3980 CCCACTCAAGTACGACTCTGGGGCACCTCTCACTGGCACCAGAAACACGACTGGGCTC 4039
Qy 1475 rLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAs 1495
Db 4040 CATCATCGGACGCCCCGGCGGCTTTCCTGCCCTGCACCCGCTGGACATAATGGCTGA 4099
Qy 1495 pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLysLysSerArgProGlyThrAl 1515
Db 4100 CGCCCGGACACTGGAGCGTGCCTGTATGAAGAGAGTCTGAAGAGCGGCTCAGGGACAG 4159
Qy 1515 aSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLy 1535
Db 4160 CAGTGTGAGGGGGCTCCATCACGCTGGGCTCCAGTCTGCTGCTGCTGAACTGGGCAA 4219
Qy 1535 sProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr 1555
Db 4220 GCCACGCAAGGCGCTTACGAAGACACACGGGGCACCCCTTACCAGTCACTCCGCTGC 4279
Qy 1555 oArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSe 1575
Db 4280 ACCTGCTCCCTGTGACACGAGGAGCGCCACGCCCTTCAGAAAGGACGCTCCT 4339
Qy 1575 rSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSe 1595
Db 4340 ATCCAGCAAGGCTCCGAGACCGGAAGCTGACATCTACACCCCGGAGATGCGCAAGTC 4399
Qy 1595 rProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLe 1615
Db 4400 CCCACACAGCACTGTGCGGAGCACCCCTCACCCTTCCCTTATGAGCACTGTCT 4459
Qy 1615 uArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProTh 1635
Db 4460 CCGGGCGGTGACGGGTGAGCTGTACCTGTGTGATCATCCCATTTGGCTTTCACCCAC 4519
Qy 1635 rSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrIleuProArgHi 1654
Db 4520 CTCCATACCCCGAGGGATCCCTCTGGAAGCAGCAGCGCAGCTACTACCTGCGCGGCA 4579
Qy 1654 sLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrPr 1674
Db 4580 CTTGGCGCCCGACCCCTTACCACCTGTATCCACCTTACTCTATCCCGGGCTACCC 4639
Qy 1674 oAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGl 1694
Db 4640 TGACACGGCGGCTTGGAGAACCGCCACCATCATCATGACTATCATCTACCTCGCAGCA 4699
Qy 1694 nMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSe 1714
Db 4700 GATGCACCAACGCTGCTCCGCCCTCCGCCCTGAGCGGTGCTGACATGCTGAGGGGTCTGTC 4759

LOCUS AF113002 7465 bp mRNA linear ROD 20-MAR-1999
DEFINITION Mus musculus silencing mediator of retinoic acid and thyroid hormone receptor beta mRNA, complete cds.
ACCESSION AF113002
VERSION AF113002.1 GI:4454549
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 7465)
REFERENCE 1 (bases 1 to 7465)
AUTHORS Ordentlich,P., Downes,M., Xie,W., Genin,A., Spinner,N.B. and Evans,R.M.
TITLE Unique forms of human and mouse nuclear receptor corepressor SMRT
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2639-2644 (1999)
MEDLINE 99178941
PUBMED 10077563
REFERENCE 2 (bases 1 to 7465)
AUTHORS Downes,M.R., Ordentlich,P. and Evans,R.M.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-1998) Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source Location/Qualifiers
1..7465
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="spleen; brain"
363..7124
/functions="transcriptional co-repressor"
/notes="mSMRT beta; longer SMRT alpha isoform encoded by GenBank Accession Number AF113001"
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid hormone receptor beta"
/protein_id="AAD20945.1"
/db_xref="GI:4454550"
/translation="MSGSTQPVQATWRAEPRYPHGI SYPVQIARSHPLYNQPSDT RQTHENIKINQMRKKLILFKRNHARKQWQFCQIDQLMEANEKVERIENNR RRAKESKREYEQPEITRKQELQERQMSRGQSLGSAASEHEVSEIIDL SEQENLEKQRLAVIPMLYDADQRIKF INNGLMDDPMKYKQDVNMWSEQR DTFRFMQHKNGLIASFLERKTVAECVLYLTCKNENYKSLVRSYRRRGKQQ QQQQQQQQQQWARSQBEKEKEKEKEKEKQDAENEKEELSKEKTDTSQED NDEKAVASGRKTANSQCRKGRITRSMANEANHEETATPQSSSLASMEWNESSRM TESEETARKKGLLEHGRNWSAIARWVGSKTVSQCKNFYFNKQRQLDELQOKLM EKERNARKKKKTPAAASEETAPPAAEDEEASGASANEELAEASQASGNEV PRVCEGSGPAVNNSSDTSVSPRSEATKDTGKPTGTEALPAATQPPVPPPEEPAV APAEPGPVDPASGPPSPESHLPHPRLILWTRMKNKPLIQLPQRMPSRSLRPRS MWKPREPRASEKPKKSDHKKETEEDKAGTEALETVSEAPLKVEKAGSKAA VTKGSSSGATQSDSSATCSADEVDREPGDKGELLSPESLLTPAGDVPASPOKP LDLUKQRAAALIPVITVKHEPPREDTVPKVPVPPVPPPTQHLQEGDVQSQQSGP RKSRSFVPAEKEAPKAPFPAPTEPGKLPTEPPRWSGGLFPPIPPREVIKTSPHA ADSASFTYPPGHPPLGLHDSARVLPRLPPI SNPPPLISSAKHPGLERQLGAIQQ GMSVOLRVSHSHAKAPMGLTWGLPVDPKLGTALGSAITKGLPSTRAADG PSYSGTTHCTADVLTKTIRIVGEDSPSLDRAREDTPKGHVIEGKKGHVLSY EGMWSVQCSKEGRSSGPPHETAPKTYDMMEGRVGTVTISAGIEGLMGRALPEQ HSPHLKEQHHRISITGIDPSRYEAGQSLREKLLKREGTTPPPPPPRDLTEYIK FRPLDPLGLKXPTHEGVTATKEAGRSIHEIPREELRTPELPLAPLAKESGSIQ GTPLKYDGSAPSTGKTHDVRSIIGSPGFPAHPHLDIMADARALACRYEESLSR SGTSSGAGGSIITGAPVPELGPQSPPLTYEDHGAPFTSHLPGRSPVTTREPTPLR QEGSLILSKASQDKLTSTPREIAKSPHSTVPEHHPHISPYEHLRGVTVGLYRGH IPIAFDTSIPRGIPLAAAAAYLPHLAPSTYPHLYPPLIRGYPDTAALENQIT IINDYITSQQMHNAASQAQADRMKGLSPRESSALNAAQPRGLIDLSQVPLPVP LVPPTGPATADRLAYTAPPPFSSSHSSPLPGGTHLAKETATSSSREER ERERDKSILSTTTTVEHAPITWRPTEQSGAGSSRPASHTHQHSISPTODALQOR PSVLNTHMKGVTSVEBPQPTVLRWARSTSTSPVRPAATFPATYHCPGLGTGVY PTLMEVLLPKETSRVARPERARVDAGHAFITKPGREPASSPSKSPESPLAPSSS HTAIARTPAKNLAPHASPDPAPTSADLHREKTSQKPFISQIELRLSLGVHSGAGY SPQGVETSPVSGFSLTHDKSKPLSEELKSHLEGRHKQFGPMKLSAAHLPHY RPLPESQSPSLQTAGIKGHQRVVTLAQHISEVITDITRHHQQLSGPLPAPLY SFFGASCPVLDLRRPPSDLYLPPDHGTGPARGPHSGGKRSPSPKTSVLGSSBAI EPVSPBEGMTEPGHARSTAYPLLYRDGEOGQERMGLESFGNTSQPTFFFSKLTESNA

ORIGIN
Alignment Scores:
Pred. No.: 2,69e-192 Length: 7465
Score: 9894.00 Matches: 1955
Percent Similarity: 79.99% Conservative: 84
Best Local Similarity: 76.70% Mismatches: 183
Query Match: 74.87% Indels: 328
DB: 10 Gaps: 30
US-09-522-753-5 (1-2517) x AF113002 (1-7465)
QY 1 MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProArgTyrPro 20
DB 363 ATGTCAAGGATCCACACAGGCTGTGGCACAGATGGCGGCTGCTGAGCCCGCTACCCA 422
QY 21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
DB 423 CCCCATGGCATCTCTTACCCGCTGCAGATAGCCCGTCCACACG----- 467
QY 41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
DB 467 ----- 467
QY 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
DB 467 ----- 467
QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
DB 467 ----- 467
QY 101 GluPheIleGluSerLysArgProArgLeuGluLeuProAspProLeuLeuArgPro 120
DB 467 ----- 467
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
DB 467 ----- 467
QY 141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
DB 467 ----- 467
QY 161 GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
DB 467 ----- 467
QY 181 ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeu 200
DB 467 ----- 467
QY 201 GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
DB 467 ----- 467
QY 221 SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
DB 467 ----- 467
QY 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
DB 468 -----CCTCTGTACAAACGACCG 485
QY 261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
DB 486 TCTGACACAGCCAGTAGTACCATGAAATACATCAAAATAACAGCGCATGCGGAAGACTG 545

1814 IleLeuThrSerThrThrValGluHisAlaProIleTrpArgProGlyThrGluGln 1833
Db ATCTTACAGTACCACTACAGTGGAGCATGCACCATCTGGAGACTGTGACGAGCAG 5532
1834 SerSerGlySerSerGlySerGlyGlyGlyGlyGlySerSerSerArgProAlaSer 1853
Db AGCAGCGGGGCT-----GGGGGAGCAGCGCGCCCGCTCC 5568
1854 HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGln 1873
Db CACACC-----CACAGCACTCGCCCATCTCCCCCGGAGCCAGAGCGCTTGACAGCAG 5622
1874 ArgProSerValLeuHisAlaThrGlyMetLysGlyIleIleThrAlaValGluProSer 1893
Db AGGCCAGTGTGTGCACAACACGAGCATCAAGGGGTGTACCTCCGTGGAACCCGGC 5682
1894 LysProThrValLeuArgSerThrSerSerSerProValArgProAlaAlaThrPhe 1913
Db AGCCCCACCGTCTGAGGTCCACTCCACTCTCTCGCCCTGTCCGCCAGCTGCCACATTC 5742
1914 ProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMet 1933
Db CACCTTGCACCCACTTGCCTTGTGGCACCCTTGAAGGGGTCTACCTTACCTCATG 5802
1934 GluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAla 1953
Db GAGCCCGTCTCTTACCCAGGAGACTCTCGGGTCCGCCGCCCGGCGCGCCGTGTG 5862
1954 AspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSer 1973
Db GACGGTGGCCATGCTTCTTCCACCAACCCCGGCCCG-----GAGCCCGCTCC 5913
1974 SerProSerLysGlySerGluProArgProLeuValProValSerGlyHisAlaThr 1993
Db TCACCCAGCAGAGCTCCGAGCCCCGATCCCTAGCACCCCCAGCTCCAGCCACAGCC 5973
1994 IleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProAla 2013
Db ATGCCCGCACCCAGCAGCAGAGCCCTTGACCCACCCACCATGCCAGTCCGAGACCCCGGG 6033
2014 ProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIle 2033
Db ---CCACCTCGGCCTCAGATCTGCACCGGAGAAAGACTCAAAGTAAACCTTTTCCATC 6090
2034 GlnGluLeuGluLeuArgSerLeuGlyTyrHis---GlySerSerTyrSerProGluGly 2052
Db CAGGAATGAATCCGCTTCTCTGGGTACACAGTGGAGCTGGCTACAGCCCCGATGGG 6150
2053 ValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeuProLys 2072
Db GTGGAGCCCATCAGCCCGGTGAGCTCCCCCAGCCTGACCCACAGCAAGGGGCTCTCCAAA 6210
2073 HisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092
Db CCTCTGGAAGAGCTAGAGAAGACCACTTGAAGAGGGAGCTGGGCACAAAGCAGCCAGGC 6270
2093 ProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSer 2112
Db CCATGAAGCTCAGCGGGAGGTGGCCATCTCCACATCTCGGCCACTGCCCGGAGAGC 6330
2113 GlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
Db CAGCCCTCATCAGCCCACTCTCCAGACTGCCCCAGGCATCAAGGTCAACAGAGGGTG 6390
2133 ValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisPro 2152
Db GTACACCTGTGCTCAGCACATCAGCAGGTATTACGAGGACTACACCCGGCACCCCGG 6450
2153 GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPro 2172
Db CAGCAGCTAGTGGGCCCTTCCGCCCTCTCTACTCTTTCGGAGGACAGCTGCCCT 6510
2173 ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAspHisGlyAla 2192

6511 GTGCTGGATCTTTCGCCGCCACCCAGTAGTACCTTACCTCCACCCCGACCATGSCACC 6570
2193 ProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThr 2212
Db CCAGCCCGGGGATCCCCCACAGTAGAGGGGGGAAAGTCTCCCAAGACCCGCAAAACA 6630
2213 SerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetThr 2232
Db TCGGTCTCTGGGAGCAGTGGAGTGCATTGAGCTGTGTCTCCACACAGAGGSCATGACT 6690
2233 GluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThr 2252
Db GAGCCAGGACATGCTCGGAGCGCTGTGTACCCACTGCTGTATCGAGACGGGAACAGGC 6750
2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
Db GAGCCC---AGATGGGCTCTAAGTCTCCAGGCAACACCCAGCGCCGCGCTTCTTTC 6807
2273 SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluLeuAsnLys 2292
Db AGTAAGCTGACTGAGAGCAACTCCGCCATGGTGAAGTCGAAGAAGCAGGAGATCAACAAG 6867
2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu 2312
Db AAACCTCAACACCAACCGGAACGAGCCAGAAATACAATATTGGCCAGGCTGGGACGAA 6927
2313 IlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db ATCTTCAACATGCCCGCCATCACTGGAGCAGGCTTATGACCTGTAGAAGCCAGCGGTG 6987
2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGly 2352
Db CAAGAACACCGCAGCACCAACATGGGCTAGAGGCCATTATTAGAAGGCACCTCATGGGT 7047
2353 LysTyrAspGlnTrpGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
Db AAATATGATCAGTGGGAGAGCGCCCGCGCTCGGGCCAAATGCTTTTAACTCTGSAAT 7107
2373 AlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
Db GCAGGCGCAGTCTGCGCCGCTGCTATGCCCATTAACCACTGCTGACGAGCGAGTGAC 7167
2392 HisThrLeuThrSerProGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
Db CACGCACTCACCTCGCCAGGTGGAGTGGGAAGCAAGGTCTCTGCGCAGACCTAGCAGC 7227
2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
Db CGAAAAGCCCAAGTCCGACAGCAGCGCTAGCGTCCGAGAGCCGACCCCTTCTGTCTCC 7287
2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGlu 2451
Db TCAGTACATCTCAGAGGGGACTGCAATCCCGGAACCACTCAACCAACGCTGTGGGAG 7347
2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
Db GACCGGCCCTCATCTGCAGGGTCCAGCCATTCCCTTACCAACCTTTGATTATGAGGCTA 7407
2472 GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491
Db CAGGAGGTGTCTATGGGCTTCCCGCCCGCCCTTGGCGCAGGACGCGGCCCTTA 7467
2492 AlaGlyProHisHisAlaTrpAspGluProLysProLeuLeuCysSerGlnTyrGlu 2511
Db GCTGGTGTCCACACCGCTGGGATGAGGAGCCCAAGCCACTGCTGTGTTCACAGATGAG 7527
2512 ThrLeuSerAspSerGlu 2517
Db ACCTCTCGACAGCGAG 7545

RESULT 5
AF113002

Db 3397 ACACCCCCGGTCAACCGCTGCTCTGGGCTCCACGATAGTCCCGGCGCCCTCTCGCCA 3456
Qy 1098 ArgProProThrIleSerAenProProProLeuIleSerSerAlaIyHisProSerVal 1117
Db 3457 CGTCCCCCCC---ATCTCTAACCCCCCACCCTCATCTCTCTCCCAAGCATCCCGCGTA 3513
Qy 1118 LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProIy 1137
Db 3514 CTTGAGAGGAGCTGGGTGCCATCTCCAGGGATGTCACTCAGCTTCGTGTGCCTCAC 3573
Qy 1138 SerGluHisAlaIyAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp 1157
Db 3574 TCAGAGCATGCCAAG---CCCATGGGCTCTCATCCATGAGCTGCGCTTGGCGTGCAC 3630
Qy 1158 ProIyIyIyLeuAlaProPheSerGlyValIyLeGlnLeuSerProArgGlyGln 1177
Db 3631 CCTAAGAAGCTG----- 3642
Qy 1178 AlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly 1197
Db 3643 -----GGG 3645
Qy 1198 ThrAlaLeuGlySerValProGlyGlySerIleThrIyLeProSerThrArgVal 1217
Db 3646 ACAGCACT-GGCTCGGCCACAGTGGAGCATCACCAAGGGCT-CCCAGTACCCGGCT 3703
Qy 1218 ProSerAspSerAlaIleThrIyArgGlySerIleThrHisGlyThrProAlaAspVal 1237
Db 3704 GCAGACGGCCCC-----AGCTACAGAGGCTCTATACCCACGC-ACGCCCGAGAGCTC 3756
Qy 1238 LeuIyIyIyGlyThrIleThrArgIleIleGlyLeuAspSerProSerArgLeuAspArg 1257
Db 3757 CTCTAAGAGTACCATCAGCAGGATCGTGGTGAGGACAGCCCAAGTCGCTTGACCG 3816
Qy 1258 GlyArgGluAspSerLeuProIyGlyHisValIleThrGlnGlyIyIyHisVal 1277
Db 3817 GCACGAGGAGACACCTTGGCCCAAGGGCCATGTCTATCATTAGGGGCAAGAAAGGCCACGTC 3876
Qy 1278 LeuSerIyGlyGlyMetSerValThrGlnCysSerIyGlyLeuAspGlyArgSerSer 1297
Db 3877 CTATCTTAAAGTGATGTCTCGTGTACAGTCTTAAGGAGGATGAAGAGGAGC 3936
Qy 1298 SerGlyProProHisGluThrAlaAlaProIySerArgThrIyAspMetGluGlyArg 1317
Db 3937 TCGGGCCCCACCCATGAGCTGCGGCCCTTAACGCACCTATTGACATGATGGAGGCGGT 3996
Qy 1318 ValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db 3997 GTAGGAGGAGCTGTCACTCAGCCAGCATAGAGGAGCTCATGGCCCGCGCATC---CCT 4053
Qy 1338 GluArgHisSerProHisHisLeuIyGlnHisIleArgGlySerIleThrGln 1357
Db 4054 GAGCAGCACAGCCCC---CACTTCAAGAGGAGCATCACATCCAGGCTCCATCAGCCAA 4110
Qy 1358 GlyIleProArgSerIyValGluAlaGlnGluAspIyLeuArgArgGluAlaIyLeu 1377
Db 4111 GGATCCCGAGGTCTATGTGGAGCGCAGAGGAGCTATTACGGCGGGAGGCCAAGCTC 4170
Qy 1378 LeuIySerGluGlyThrProProProProProProSerArgAspLeuThrGluAlaIy 1397
Db 4171 TTGAAGCGAGAAGGAGACACACCTTCCGCCACACCACTCTGGGACCTGACTGAGACCTAC 4230
Qy 1398 LysThrGln-----AlaLeuGlyProLeuIySerLeuIyProAlaHisGluGlyLeu 1414
Db 4231 AAGCCCCGGCCCTGGACCTCTGGGTCCTTGAAGCTGAAGCGGCTCAGAGGGTGTG 4290
Qy 1415 ValAlaThrValIyGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434
Db 4291 GTACCAACTGTGAAGGAGGCGGCGCTCTATCCATGAGATCCCGAGAGGAGCTGCGC 4350
Qy 1435 HisThrProGluLeuProLeuAlaProArgProLeuIySerIleThrGlnGly 1454
Db 4351 CGCACCTGAGCTACCCCTGGCACCCACCGGCTCTGAAGAGGGGTTCATCACCCAGGGC 4410

Qy 1455 ThrProLeuIySerIyAspThrGlyAlaSerThrThrGlySerIyIyHisAspValArg 1474
Db 4411 ACCCACTCAAGTACGACTCTGGGCACCTCCACTGGCACCAAGAAACACGACGTGGC 4470
Qy 1475 SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494
Db 4471 TCATCATCGGAGCCCGCGCGCTTTCCTCTGCCCTGCACCCGCTGACATAATATGGCT 4530
Qy 1495 AspAlaArgAlaLeuGluArgAlaCysIyGluGluSerLeuIySerArgProGlyThr 1514
Db 4531 GACGCCCGGGCACTGAGCGTGCCTCTATGAAGAGAGTCTGAAGAGCGGTCAGGAGACC 4590
Qy 1515 AlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGly 1534
Db 4591 AGCAGTGGTGCAGGGGCTCCATCACACGTGGGCTCCAGTCTGCTGCTGAACTGGGC 4650
Qy 1535 LysProArgGlnSerProLeuThrIyGluAspHisGlyAlaProPheAlaGlyHisLeu 1554
Db 4651 AAGCCACGGCAAGCCCACTGACTTACGAAGACACCGGGGCACCTTCCACGATCACCTG 4710
Qy 1555 ProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeu 1574
Db 4711 CCAGTGGCTTCCCTGTGACCAAGAGGAGCCACCGCACCTTCCAGGAAGCAGCTC 4770
Qy 1575 SerSerSerIyAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys 1594
Db 4771 CTATCCAGCAAGGGCTCCCGAGGACCGGAAGCTGACATCTACACCCGGGAGATCGCAAG 4830
Qy 1595 SerProHisSerThrValProGluHisHisProHisProIleSerProIyGluHisLeu 1614
Db 4831 TCCCCACACAGCACTGTGCCCGAGCACCACTTCCACCCATCTCCCTTATGAGCACTTG 4890
Qy 1615 LeuArgGlyValSerGlyValAspLeuIyArgSerHisIleProLeuAlaPheAspPro 1634
Db 4891 CTCGGGGGCTGACTGGTGTGGACCTGTACCGTGTGCATCCATTCCTTGGCTTTGACCC 4950
Qy 1635 ThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaIaIyIyIyLeuProArg 1653
Db 4951 ACCTCCATACCCCGAGGATCCCTCTGGAAGACGAGCAGCGCTACTTACCTGCGCCGG 5010
Qy 1654 HisLeuAlaProAsnProThrIyProHisLeuIyProProIyIleuIleArgGlyIy 1673
Db 5011 CACTTGGCCCCAGCCGCCACTACCCACACCTGTACCCACCTTACCTCATCCCGCGCTAC 5070
Qy 1674 ProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspIyIleThrSerGln 1693
Db 5071 CCTGACAGCGCGCCCTGGAGAACCCGACAGCATCATCATGACTTACATCATCCTCGCAG 5130
Qy 1694 GlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeu 1713
Db 5131 CAGATGACCAACAACGCTGCTCCGCCATGGCCCGAGCGTGTGACATGCTGAGGGGTCTG 5190
Qy 1714 SerProArgGluSerSerLeuAlaLeuAsnIyAlaAlaGlyProArgGlyIleIleAsp 1733
Db 5191 TCACCGGAGAGTCTCGCTGGGCTCAATTAATTCGCTGGCCCAAGAGGAGCATTTACGAC 5250
Qy 1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
Db 5251 CTGTCCCAAGTGCACACCTGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5310
Qy 1754 AlaMetAspArgLeuAlaIyLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Db 5311 GCCATCGACCGCTTGCCTACCTCCCACTCCCGCTGCGCCGCCACCTTCCAGCAGCGCCACAGT 5370
Qy 1774 SerSerProLeuSerProGlyGlyProThrHisLeuThrIyIyProThrThrThrSerSer 1793
Db 5371 AGCTCACCGCTGTCCCAGGAGGCCCACTCACCTAGCTAAACCACTGCGCATCTTCA 5430
Qy 1794 SerGluArgGluArgAspArgGluArgAspArgGluArgGluArgGluArgGluArgGluArg 1813
Db 5431 TCGGAGCGGGAACGGGAACGTGAGCGGGAACGAGAC-----AAGTCC 5472

source	1. .8544	/organism="Mus musculus"	21	ProHisserLeuSerTyProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu	40
		/mol_type="mRNA"	220	CCCCATGGCATCTCTACCGGTGAGTAGCCCGGTCCACACGAGCGTGGGCTGCTT	279
gene	1. .8544	/db_xref="taxon:10090"	41	GlutylGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleleGln	60
CDS	160. .7548	/gene="Smrte"	280	GAGTACCAACACACCCCGTGTACTACCTCACACCTGTCCCGGTTCATCATCCAG	339
		/gene="Smrte"	61	ProGlnArgArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGln	80
		/function="transcriptional corepressor"	340	CCACAGAGAGCGGCCCTCATCTGTGTGAGAGTTCACGCTGGAGTGAACGGTCTCAG	399
		/codon_start=1	81	GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet	100
		/product="silencing mediator of retinoic acid and thyroid hormone receptor extended isoform"	400	GAGCTCCACCTGGCCCTGAGTCCCGCAGCTTCTGCTGAGCTGGGCAAGCCGACATA	459
		/protein_id="AAD22972.1"	101	GluPheIleGluSerIleArgProArgLeuGluLeuLeuProAspProIleuLeuArgPro	120
		/db_xref="GI:4559296"	460	GAATTCACCGAGAGCAAGCGCCCGCTGGAGCTACTACCCGATACCTGCTGCGCCCA	519
		/translation="MSGSTQPVQAOTWRAEPYPHGISYFVQIARSHTDVGLLEVOH	121	SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer	140
		HPDYTHSLSPGSIQIQRRLPSELLSEFPQGSERSQELHLPESRTFLPELKGDELIF	520	TCACCCCTCTGCGCACTGCGCAGCGAGTGGTCTGAAGACCTTACCAAGGACCGTAGC	579
		TSKRPLLEPDLTLRPLPGLATQSGSEDITKDRSLAGLKEPVPSPHPADPEL	141	LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu	160
		ELAPSLKEELITQNRDLVDREITWVEQIISKLKKQQLQEEAAKPEPEKVPSP	580	CTGGCAGGCAAGCTGGAGCTGTGTTCACCTCCAGTCCCGCAGCTGACCTGAGCTA	639
		PLSKRSILVQIITDENRKAEEAHRILEGLPOVELPYNQPSDTROYHENIKINQA	161	GluLeuValProProArgLeuSerIleGluLeuLeuIleGlnAsn---MetAspArgVal	179
		MRKLLYFKRRNHARKQWQRCQRYDOLMEAWKKVERIENNRPRRAKESKVRYY	640	GAGTGGCCCACTCTCGAGTGTCCAGAGGAGAGTGTATCCAGAACAGATTTGAGCCCGTG	699
		EKQPFIRQRELQRMQRVQSGGLSMAARSEHYSEIIDGLSELENLEKMRQ	180	AspArgGluIleThrMetValGluGlnGlnIleSerLysLysLysLysGlnGlnGln	199
		LAVIRHVRDQRIKFINNGLMDPMKIVQVINWMSQEDITFREKFMQHPKQ	700	GACCGTGAGATACCATGTGTAGCAGCAGATCTCCAGCTGAGAGAGAGCAGCAACAG	759
		FGLIASFLEKTVAECLVLYLTKNENKSLVRSYRRGSKSQOQQOQQOQQOQQ	200	LeuGluGluGluAlaLysProGluProGluLysProValSerProProIle	219
		ARSQBEKEKEKEKEKEKEKEKEKEKEKEKEDTDTSGEDNKEAVASKGR	760	TTGGAGGAGGAGCGCCCAAGCGCCGCAACCCGAGAGCTGTGTGCGCCACCCACATA	819
		TANSQGRRIITRSMANANHEETATPOQSELSAMENNESRWTEEMETAKGILL	220	GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGlu	239
		EHCNRSALTARVGSITVQCKNFYFNKKRONLDEILOQHLKMEKERNARRKKKT	820	GAATCAAGACCGAAGACCTGTCTCAGATCATCTACGATGAGAACCGGAGAAAGCCGAA	879
		PAASSETAPPAADEEMASASANEELAEAEASQASNEVPRVCEGCPAAVN	240	AlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGln	259
		NSSDTSVSPSPSEATQDQKPTGTEALPAATQPPVPPPEPAAAPABSPVPDASG	880	GCCGCAACCGGATCTTAGAGGCTTGGGCGCCCGAGTGGAGTGGCTCTGTATCAACAG	939
		PPSPSPSPAPPAATVDQDEAPAPAPQTEDEAKESEAEIDVKGPEPSEAE	260	ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLys	279
		PPSVSKDHEEPEEKAGTEAETVSEAPLKEVEAGKAAVTGKSSSGATODS	280	LeuIleLeuTyrPheLysArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCys	299
		DSATSCADEVDPEGGDKRLRPLSLITPAGDPRASTPOKPLDLKQLKORAAAI	1000	CTGATCTTGTACTTTAAGCGGAGAACCCAGCGCGCAAGCAGTGGGAAACAGCCCTCTGC	1059
		PPITVKEHPREDTVPKVPVPPPTOHLQPEGDVQSGSGPRGSKRSVPVPAEK	300	GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsn	319
		EAKPAFFAPFTEGQSYRLSPHAGHLRSHHPREVITKSTRADPLFSTVPGHPLPL	1060	CAGCGTATGACCGAGTCTATGGAGGCTGGGAGAGAGGTAGCGCATAGAACAAAT	1119
		GLHDSARPLPRPISNPPLIISAAKHGVLRLQGLAISQGMVQLRVHSEHAKPMG	320	ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu	339
		PLTMEPLAVDPKLLGTALAPPVEASPRASQYPCRRPQLQLXHPHPTADVLKGT	1120	CCGCAAGAGGAGGCGCAAGGAGCAAGGTGAGGAGTACTTACGAGAAACAGATTCCCGGAG	1179
		ISRIVGDSRLDRAREDTLPKGVHYEGKGHVLVSYEGGMSVSCSKEDRSSGP	340	IleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlnGlnArgGlySer	359
		PHETAAPKRTYDMWGRVGTVTASIEGLMGRAIPEQHSPLKEQHIRGISTQIP	1180	ATCCGCAAGCAGCGGAGCTGCGAGGAGCGCATCAGACGAGGTGGCCACGCTGGCAGT	1239
		RSVTEAQEDYLREAKLKEGTPPPPPDRDLTETYPKPLDPLGLPLKPTHEGV	360	GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleleAspGlyLeu	379
		ATVKEAGRSIHETPRELRTPELPLAPRLKEGSIQTGTPLKYDSGAPSTGTXKHIV	1240	GGGCTCTCCATGTTCGGCTGCCCGACGTGAGCATGAGGTTTCTGAGATCATTTGATGGCTTG	1299
		RSIIGSPFPFPAHPLDMDADALERACTEESLSRSGTSSGAGSITRGAAPVVP			
		ELGKQPSLTYDHDGAPTHLPRGSPVTTRETPRLQESLSSKASQDRKLTTTP			
		REIAPKPSHTVSHHPTSPYHLIRGVTGDLVYRCHPLAFDPTSIPRGPLEASA			
		AAVYLRHLAPSTPHLYPYLIRGVPDTAALENQTIINDYTSIQMHNAASAMA			
		QADMLRGLSPRESSIALNYSAGPRGIDLSQVPLPLVPPPTGPTATADRLAYLP			
		TAPPPSSHSSSPLSPGTHLAKPTATSSSEREREEREDKSLTSTVTEHAPI			
		WRFGTQSSGAGSSPASHQHSFISPTQALQQRSPVLNMTSMKGVTSVERGT			
		PTVLRSTSSPVVRPAATFPFATHCPGLGTLGVYPTLMEPVLPLPKETSRKVARPR			
		VDGHAFLTKPPAREPASSPKSEPRSLAPPSSHTAARTPAKSLAPHASPDPPG			
		PTSASDLHREKTKSPFSIOELRLSLGVSHGAGYSPDVEPTSPVSSPSLTHADKGLS			
		KPLELEKSHLEGLRHKQPMKLSAEAAHLPLRLPESQSPSPQLQTAPGKIGL			
		QRVVTLAQHTSEVITQDYTHHPQQLSGPLPAPLYSPFGASCPLDLRPPSLYLGP			
		PDGHTFARGSPGEGKSPKSPKTSVLGSSSEDAIEPVPPEGWTPEGHARSAYPLL			
		YRDGQEPFPMGSKSPNTSQPAFFSKLSTESNAMSVMKSKKINKKLNTHNRPEY			
		NIQCGTEIFNMPAITGAGLMTCSQAVQSHASTNMGLEAIRKALMGYDQWEEPP			
		LGNANFLNLSALPAAMPTITADRSQHALTSPCGGKAKVSRGKSDYKESAPP			
		GLASGDRPPSVSVSHSGDCNRRITPLTNRVEDRPPSAGSTPPYPLNMLRLQAVMA			
		SPPPGLAAGSGPLAGPHHAWDEBPCLLCQYETILSDSE"			

ORIGIN

Alignment Scores:

Pred. No.:	3,116-211	Length:	8544
Score:	10832.50	Matches:	2149
Percent Similarity:	87.94%	Conservative:	90
Best Local Similarity:	84.41%	Mismatches:	197
Query Match:	81.97%	Indels:	117
DB:	10	Gaps:	33

US-09-522-753-5 (1-2517) x AF125671 (1-8544)

Qy	1	MetSerGlySerThrGlnLeuValAlaGlnThrTyrArgAlaThrGluProArgTyrPro	20
Db	160	ATGTGAGGATCCACACAGCTGTGGCAGACATGCGCGGCTGCTGAGACCCCGCTACCCA	219

```

Db      6292  TACCCCTCATGGAGCCGCTGTTTACCAAGGAGACCTCTCGGGTCGCCGCCGCGAGCG 6351
Qy      1950  gProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuG1 1970
Db      6352  GGCCTGGGTGGAGCGTGGCCATGCCCTTTTACCAAAACCCCGGGCGG-----GA 6402
Qy      1970  uProAlaSerSerProSerLysGlySerGluProArgProLeuValProValSerG1 1990
Db      6403  GCGCGCTCTCACCCAGCAAGAGCTCCGAGCCCGCATCCCTAGCACCCCGCTCCAG 6462
Qy      1990  yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAs 2010
Db      6463  CCACACAGCATCGCCCGCACCCAGCAAGAACCTTGACCCACCATGCCAGTCCGGA 6522
Qy      2010  pProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPr 2030
Db      6523  CCGCCCGGG-----CCACCTCGGCCTCAGATGTGCACCGAGAAAGACTCAAGTNAACC 6579
Qy      2030  oPheSerIleGlnGluLeuLysArgSerLeuGlyTyrHis---GlySerSerTyrSe 2049
Db      6580  CTTTTCCATCCAGGAATTGGAATCTCGTTCTCTGGGTACACAGTGGAGCTGGCTACAG 6639
Qy      2049  rProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysG1 2069
Db      6640  CCGCATGGGTGGAGCCCATCAGCCCGGTGAGCTCCCGCAGCTGACCCACGACAGGG 6699
Qy      2069  yLeuProLysHisLeuGluLeuLysAspLysSerHisLeuGluGlyLysLeuArgProLy 2089
Db      6700  GCTCTCCAAACCTCTGGAGAGCTAGAGAGAGCCACTTGGAGGGGAGCTGGCGACAA 6759
Qy      2089  sGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLe 2109
Db      6760  GCAGCCAGCCCATGAGCTCAGCGCGGAGGCTGCCCATCTCCACATCTGCGGCCACT 6819
Qy      2109  uProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHi 2129
Db      6820  GCCCGAGAGCAGCCCTCATCCAGCCCATCTCTCCAGACTGCGCCCGAGGATCAAGGTCA 6879
Qy      2129  sGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrAr 2149
Db      6880  CCAGAGGTGTGTCACCTGCTCAGCACATCAGCAGAGTCAITTACGACGAGCTACACGG 6939
Qy      2149  gHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAl 2169
Db      6940  CCACACCCGCGCAGCTCAGTGGCCCTTCCCGCCCTCTCTACTCTCTCTCTCTCTCTCT 6999
Qy      2169  aSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAs 2189
Db      7000  CAGCTGCCCTGTCTGTGATCTTCGCGCCCGCCAGCTGACCTTACTCTCCACCCCGCA 7059
Qy      2189  pHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluPr 2209
Db      7060  CCATGGCACCCAGCCCGGGATCCCCCACTGAGGGGGCAAGAGTCCCAAGACC 7119
Qy      2209  oAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProProG1 2229
Db      7120  CACCAAAACATCGTCTCGGCAGCAGCAGGAGTGCATTTAGCCCTGTGTGCCACCA 7179
Qy      2229  uGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspG1 2249
Db      7180  GGGCATGACTAGCCAGGACATGCTCGGAGCACTGGGTATCCCACTGCTGTATCGAGCG 7239
Qy      2249  yGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProPr 2269
Db      7240  GGAACAGGGCGAGCCCC---AGGATGGTCTAGAGTCTCCAGGCAACACAGCCGCGCC 7296
Qy      2269  oAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnG1 2289
Db      7297  AACCTTCTTCAGTAAGCTGACTGAGAGCACTCGGCATGCTGAGTGAAGTCGAGAGCAG 7356
Qy      2289  uIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnPr 2309

```

```

Db      7357  GATCAACAAGAAACTCAACACCAACCGGAACGAGCCAGAAATACAATATTGGCCAGCC 7416
Qy      2309  oGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSe 2329
Db      7417  TGGAGCGGAATCTTCAACATGCCGCCATCACTGGAGCAGGCTTATGACCTGTAGAAG 7476
Qy      2329  rGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAl 2349
Db      7477  CCAGCGGTGCAAGAACACGCCAGCAGCAACATGGGGCTAGAGCGCATTTATTAGAAAGC 7536
Qy      2349  aLeuMetGlyLysTyrAspGlnTrpGluLysSerProProLeuSerAlaAsnAlaPheAs 2369
Db      7537  ACTCATGGTAATATGATCAGTGGAGAGAGCCCGCCGCTCGGCGCCAAATGCTTTTAA 7596
Qy      2369  nProLeuAsnAlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspG1 2388
Db      7597  CCCTCTGAATCCAGCGCCAGTCTGCCGCTGCTGCTATGCCCATTAACCACTGCTGACGG 7656
Qy      2388  yArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyAr 2408
Db      7657  ACGGAGTACACCGCACCTCACCTCGCCAGTGGAGTGGGAAAGCCCAAGGTCTCTGCGCAG 7716
Qy      2408  gProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProPr 2428
Db      7717  ACTAGCAGCGAAAGCCAGTCGCCAGCACAGGCTTAGCGTCCGGAGACCGACCCCC 7776
Qy      2428  oSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnAr 2448
Db      7777  TTCTGTCTCTCAGTACACTCAGAGGGGAGCTGCAATCGCGACACCACTCACCAACCG 7836
Qy      2448  gValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIl 2468
Db      7837  TGTGTGGAGGAGCCGCGCTCATCTGCGGGTCCAGGCTTCCCTACACCCCTTTGAT 7896
Qy      2468  eMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProAlaGlySe 2488
Db      7897  TATGAGGCTACAGGAGGAGTGTATGGCTTCCCGCCCGCCCGCTTGGCGAGGCGAG 7956
Qy      2488  rGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysProLeuLeuCysSe 2508
Db      7957  CCGGCCCTTAGTGTGTCTCCCGCCAGCGCTGGATGAGAGGCCCAAGCCACTGCTGTGTT 8016
Qy      2508  rGlnTyrGluThrLeuSerAspSerGlu 2517
Db      8017  ACAGTATGAGACACTCTCGACAGCGAG 8044

RESULT 4
AF125671
LOCUS
DEFINITION
Mus musculus silencing mediator of retinoic acid and thyroid hormone receptor extended isoform (Smrte) mRNA, complete cds.
ACCESSION
AF125671
VERSION
AF125671.1
KEYWORDS
GI:4559295
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 8544)
Park,E.J., Schroen,D.J., Yang,M., Li,H., Li,L. and Chen,J.D.
SMRte, a silencing mediator for retinoid and thyroid hormone receptors-extended isoform that is more related to the nuclear receptor corepressor
Proc. Natl. Acad. Sci. U.S.A. 96 (7), 3519-3524 (1999)
JOURNAL
MEDLINE
99199215
PUBMED
10097068
REFERENCE
2 (bases 1 to 8544)
Chen,J.D.
Direct Submission
AUTHORS
Submitted (03-FEB-1999) Pharmacology and Molecular Toxicology, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA
JOURNAL
Location/Qualifiers
FEATURES

```

QY	1217	lProSerAspSerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspVa	1237
DB	4192	GCAGAGCGGCCCC-----AGCTACAGAGGCTCTATCATCCACGCGACGCCGACAGCT	4245
QY	1237	lLeuTyArgGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspAr	1257
DB	4246	CCTCTACAAGGGTACCATTACAGCAGGATCGTCGGTGGAGCAGCGCCAAAGTCGCTTGACCG	4305
QY	1257	gGlyArgGluAspSerLeuProLysGlyHisValIleTyArgGlyLysGlyHisVa	1277
DB	4306	GGCAGAGAGAGACACCCCTGCCCAAGGCCATGTCATCTATCAGGGCAGAGAAAGCCACGCT	4365
QY	1277	lLeuSerTyArgGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSe	1297
DB	4366	CCTATCTCTAGAGGTGGTATGTCGCGTGCACAGCTGCTCTTAAGGAGATCGAAGGAGCAG	4425
QY	1297	rSerGlyProProHisGluThrAlaAlaProLysArgThrTyArgMetMetGluGlyAr	1317
DB	4426	CTCGGGGCCCAACCCCATGAGACTCGCGCCCTAAACGCACCTATGATGATGAGGGGCGG	4485
QY	1317	gValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPr	1337
DB	4486	TGTAGGACAGACTGTCACTCAGCCAGCATAGAGGACTCATGCGCGCGGCCATC---CC	4542
QY	1337	oGluArgHisSerProHisLysLeuLysGluGlnHisIleArgGlySerIleThrGl	1357
DB	4543	TGAGCAGCAGAGCCCC---CACCTCAAGGAGCAGCATCATCCGAGGCTTCCATCAGCA	4599
QY	1357	nGlyLeProArgSerTyValGluAlaGlnGluAspTyLeuArgArgGluAlaLysLe	1377
DB	4600	AGGCATCCCGAGGTCTATGTGGAGGCGCAGAGGACTACTTACGCGGAGAGGCCAAGCT	4659
QY	1377	uLeuLysArgGluGlyThrProProProProProSerArgAspLeuThrGluAlaTy	1397
DB	4660	CTTGAAGCAGAGAGGGACACACCACCCCCACACCACTCGGACCTGACTGAGACCTA	4719
QY	1397	rLysThrGln-----AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLe	1414
DB	4720	CAAGCCCCGGGCCCTGAGACCTCTGGGTGCCCCCTGAAGCTGAAGCCGACTCACGAGG	4779
QY	1414	uValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuAr	1434
DB	4780	GGTAGCAACTGTGAAGGAGGCGGCGCTCTATCATGAGATCCGAGAGAGAGCTGCG	4839
QY	1434	gHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGl	1454
DB	4840	CCGCACACCTGAGCTACCCCTGGCACACAGCGCTCTGAAGGAGGGTTCCATCACCCAGG	4899
QY	1454	yThrProLeuLysTyArgThrGlyAlaSerThrThrGlySerLysLysHisaspValAr	1474
DB	4900	CACCCCACTCAAGTACGACTCTGGGGCACCCCTCCACTGGCACCAAGAAACACGACGTGG	4959
QY	1474	gSerLeuIleGlySerProGlyArgThrPheProProVaHisProLeuAspValMetAl	1494
DB	4960	CTCCATCATCGGCAGCCCGCGCGCTTTCCTTCCCTGCGCTGCACCGCTGGACATAATGCG	5019
QY	1494	aAspAlaArgAlaLeuGluArqAlaCysTyArgGluLysSerLeuLysSerArgProGlyTh	1514
DB	5020	TGACGCCCGGGCATCTGAGCGTGCCTGCTATGAAGAGAGTCTGAAGAGCCGCTCAGGAC	5079
QY	1514	rAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGl	1534
DB	5080	CAGCAGTGGTCAGGGGCTCCATCACACAGCTGGGGCTCCAGTCGTGCTGAACTGGG	5139
QY	1534	yLysProArgGlnSerProLeuThrTyArgLysHisGlyAlaProPheAlaGlyHisLe	1554
DB	5140	CAAGCCACGGCAAAAGCCCACTGACTTACGAAGACCCAGGGGACCCCTTCCACAGTCACT	5199
QY	1554	uProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLe	1574
DB	5200	GCCAGTGGTCCCCCTGTGACCAAGAGGGAGCCCCACGCCACCGCTTTCAGGAAGGACGCT	5259

Qy		1574	uSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaVal	1599
Dd		5260	CCTATCCAGCAAGCGCTCCAGACCGGAAGTGCATCTACACCCTGGGAGATGCCCAA	5319
Qy		1594	sSerProHisSerThrValProGluHiShisProHisProIleSerProTyrgluHisle	1614
Dd		5320	GTCCTCCACACAGCATGTGTGCCGAGCACCACTCTCACCCTTGATGACACTT	5379
Qy		1614	lUeAArgGLyVAlserGLyValAspLeuTYrArGserHisIlleProLeuAlaPheaspPr	1634
Dd		5380	GCTCCGGGGGCTGACTGTGTGTGACCTGTACCTGTGTGCATCTCCATTGGCTTGACCC	5439
Qy		1634	oThrSerIleProArgLyIleProLeuASP-----AlaAlaLaAlaTYrTYrlenProAr	1653
Dd		5440	CACCTCCATAACCCGAGGGATCCCTCTGGAAGCAGCAGCGCAGCGCTACTACCTGCCCG	5499
Qy		1653	gHisLeuAlaProASNProThrTYrProHisileuTYrProTYrLeulearGLYTY	1673
Dd		5500	GCACTTTGGCCCCCAGAGCCCACCTTACCACCACTGTATTACCACCTTACTCATCCGCGCGCTA	5559
Qy		1673	rProASpThrAlaAlaleuGluAsnArgGlnThrIIelleAsnAspTYrIleThrSerGl	1693
Dd		5560	CCCTGCACACGGCGGCTCTGGAGAACCAGCCACCATCATCATGATCATCATCACCTCGCA	5619
Qy		1693	nGInMetHisIAsnThralaThralaMetAlaGlnArGaAlaASPmetIeuARGLYLe	1713
Dd		5620	GCAGATGCACCAACAACGCTGCCTCCGCCATGGCCCAAGCGTGTGCATGATGTGAGGGGTCT	5679
Qy		1713	uSerProARglusErSerleuAlaleuanTYrAlaalGLyProARglylelleleas	1733
Dd		5680	GTCACCCGAGAGTCTCTCGCTGGCCCTCAATTATGCCGTGGCCCAAGAAGGCATTATCGA	5739
Qy		1733	pLeuSerGlnVALproHisleuProValleuValProProThrProGLYThrProAlaTh	1753
Dd		5740	CCGTGTCCAAGTGCCACACCTGCCGCTGTGTGTGGCCACCAACGCCAGGACCCCTGCCAC	5799
Qy		1753	rAlaMetASPArgLeuAlatyrLeuProthrAlaProGlnProPheserSeraRghIsse	1773
Dd		5800	CGCATCGACCGGCTTGCCTTACCTGCCACTGCCGCCCCAACCCCTTCAGACGCGCCACAG	5859
Qy		1773	rSerSerProLeuSerProGLYtyrothrHisLeuthrLysProthrThrThrSerSe	1793
Dd		5860	TAGCTACACCGCTGTCCCAAGAGGCCCACTACCTAGCTAAACCAACTGCCACATCTTC	5919
Qy		1793	rSerGluARgluARgaSPArgAgSLuARgaSPArgSLuARgluLylusse	1813
Dd		5920	ATCGAGCGGGGALCGGAACGTAAGCTGAGCGGGAACGAGAC-----AAGTC	5961
Qy		1813	rIleLeuThrSerThrThrValGluHiSaAlaProIleTrpARgProGLYThrGluGI	1833
Dd		5962	CATCTTCACGCTACCACTACGTGGAGCATGACCCATCTGGAGACCTGGTAGCGAGCA	6021
Qy		1833	nSerSerGLYSerSerGLYSerSerGLYGlyGLYGlySerSerSerArgProAlase	1853
Dd		6022	GAGCAGCGGGGCT-----GGGGGGCAGCAGCGCCCCCGCTC	6057
Qy		1853	rHisSerHisalAhISGlnHisSerProIleSerProAtqThrGlnAspAlaLeuGlnGI	1873
Dd		6058	CCACACC-----CACCAGCAGCTGCCCATCTCCCCCGAGACCAGGACGCTTGGAGCA	6111
Qy		1873	nArgProSerValleuHisasnThrGLYMetLysGLYIleIleThraLaValGluProse	1893
Dd		6112	GAGGCCAGTGTGTGTGCACAAACAGCAGCATGAAGGGCGTGTTCACCTCCGTGGAACCCCG	6171
Qy		1893	rLysProThrvalluu-----ArgSerThrSerThrsErSerProvalargProAl	1910
Dd		6172	CACGCCCAACGGTCTCGAGGTGGGCGAGGTCCACCTCCACCTCTTCGCTGTCCGCCGACG	6231
Qy		1910	aAlaThrPhaeProAlathrhHisCySProLeuGLYThrLeuAspGLYValTYrPr	1930
Dd		6232	TGCCACATPTCCACTGCCACCACTGCCCACTTGTGTGGCACCTTGTGAAGGGGTCTACCC	6291
Qy		1930	oThrLeuMetGluProValleuLeuProLysGluAlaProArgValAlaargProGluAr	1950

Db 6282 GTGGAGCCCGTGTGAGCTCACCAGTCTGACCCACGACCAAGGGGCTCCCCAA 6341
Qy 2072 sHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProG1 2092
Db 6342 GCACCTGGAAGAGCTCGACACAGAGCCACCTGGAGGGGAGCTCGGCCACAGCAGCCAGG 6401
Qy 2092 yProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLysGlnProGluSe 2112
Db 6402 CCCCGTGAAGCTTGGGGGAGGCGCGCCACCTCCACACCTCGCGCGCTGCTGAGAG 6461
Qy 2112 rGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVa 2132
Db 6462 CCAGCCCTCGTCCAGCGCGCTGTCCAGACCGCCCGAGGGGTCAAGGTCAACAGCGGGT 6521
Qy 2132 lValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisIlePr 2152
Db 6522 GGTCAACCTGGCCAGCACATCATGTGAGTTCATCACAGGACTACACCGGCACCAACC 6581
Qy 2152 oGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysPr 2172
Db 6582 ACAGCAGCTCAGGCGACCCCTGCGCGCCGCCCTCTACTCTCTCTGCGGCAGCTGCC 6641
Qy 2172 oValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAspHisGlyAl 2192
Db 6642 CGTCCTGGACCTCGCGCGCCACCCAGTGACCTTACTCTCCCGCCCGGACCATGGTGC 6701
Qy 2192 aProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysTh 2212
Db 6702 CCCGCGCGTGGTCTCCCCCAGCAGCAGGAGGGGCGAGAGTCTCCAGAGCCAAACAGAC 6761
Qy 2212 rSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGlyMetTh 2232
Db 6762 GTCGGTCTTGGTGGTGGTGGAGCAGCGATTGAACCTGTGTCCACCGGAGGCGATGAC 6821
Qy 2232 rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh 2252
Db 6822 GGAGCCAGGCGACTCCCGGAGTCTGTGTACCGCGTGTGTACCGGATGGGAACAGAC 6881
Qy 2252 rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh 2272
Db 6882 GGAGCCAGCAGAGATGGGCTCCAGTCTCCAGCGACACACAGCAGCGCGCCAGCTTCTT 6941
Qy 2272 eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGlnLysLeu 2292
Db 6942 CAGCAAGCTCAGCAGAGCAACTCCGCGCATGGTCAAGTCCAGAGCAAGAGAGATCAACAA 7001
Qy 2292 sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGl 2312
Db 7002 GAAGCTGAACACCCACACCGGAATGAGCCTGAATACATATATATATATATATATATAT 7061
Qy 2312 uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa 2332
Db 7062 GATCTTCAATATGCGCCCATCACCGGACAGCGCTTATGACCTATAGAGCAGCGCGGT 7121
Qy 2332 lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGl 2352
Db 7122 GCAGGAACATGCCAGCACCAATGGCGTGGAGGCCATAATATAGAAAGGCACTCATGG 7181
Qy 2352 yLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAs 2372
Db 7182 TAAATATGACAGTGGGAAGAGTCCCGCGCGCTCAGCGCCAAATGCTTTTAAACCTCTGAA 7241
Qy 2372 nAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHi 2392
Db 7242 TGCAGTGCCAGCGCTGCGCCCTGCTATGCCATTAACCGCTGTGACGAGCGAGTGAACCA 7301
Qy 2392 sThrLeuThrSerProGlyGlyGlyLysAlaLysValSerLysGlyArgProSerSerAr 2412
Db 7302 CACACTCACCTCGCCAGTGGCGGCGGAAGGCCAAGGTCTCTGGCAGACCCAGCGCG 7361
Qy 2412 gLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSe 2432
Db 7362 AAAAGCCAGTCCCGCGCGCTGGCATCTGGGAGCCCGGCGCCACCTCTCTCTCTCTCT 7421

Qy 2432 rValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAs 2452
Db 7422 AGTGCACTCGGAGGAGAGCTGCAACCGCGGAGCGCTCACCAACCGCGTGTGGGAGA 7481
Qy 2452 pArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGl 2472
Db 7482 CAGGCGCTCTGTCGAGGTTCCACGCCATTCCCTACAAACCCCTGATCATGCGGTGCA 7541
Qy 2472 nAlaGlyValMetAlaSerProProGlyLeuProAlaGlySerGlyProLeuAl 2492
Db 7542 GCGGGTGTATGCTTCCACCCCGCGGCTCCCGCGGAGCGGCGCCCTCGC 7601
Qy 2492 aGlyProHisHisAlaTrpAspGluGlyProLysProLeuLeuCysSerGlnTyrGluTh 2512
Db 7602 TGGCCCCCACCACGCTGGGAGGAGGAGCCCAAGCCACTGCTCTGCTGCAGTACGAGAC 7661
Qy 2512 rLeuSerAspSerGlu 2517
Db 7662 ACTCTCCGACGAGGAG 7677
RESULT 3
AF113001
LOCUS
DEFINITION
Mus musculus silencing mediator of retinoic acid and thyroid hormone receptor alpha mRNA, complete cds.
ACCESSION
AF113001
VERSION
AF113001.1
KEYWORDS
GI:4454547
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 8388)
AUTHORS
Ordentlich,P., Downes,M., Xie,W., Genin,A., Spinner,N.B. and Evans,R.M.
TITLE
Unique forms of human and mouse nuclear receptor corepressor SMRT
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2639-2644 (1999)
MEDLINE
99178941
PUBMED
10077563
REFERENCE
2 (bases 1 to 8388)
AUTHORS
Downes,M.R., Ordentlich,P. and Evans,R.M.
TITLE
Direct Submission
JOURNAL
Submitted (11-DEC-1998) Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
Location/Qualifiers
1. .8388
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="spleen; brain"
626. .8047
/function="transcriptional co-repressor"
/note="MSMT alpha"
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid hormone receptor alpha"
/protein_id="AAD20944.1"
/translation="MSGSTQPVATQWRAEAPRYPHGISYVPQVIARSHDTDVLLEYQH
HPTDVTSHLSGSIIPQRRRLPSLSEFQSGSERSOELHLPESRTFLPLGKPDLEF
TSKPRLELLPTLLRPSPLATQPSGSEDLTKDRSLAGLKPVSPPSPHADPEL
ELASRLSKEELIQNDVRDEITWVEQIISLKKQQLEREAAKPPPEKVPSP
ISKRSLSVQIYIDENRKAEEAHLLEGLPQVELPLYNQSDTFQYHENIKINQAM
RKXLLYIKRNRHARKQWRFQRYDLMWEAKKVERIENNRRAKESKREYIE
KQFPIRKQRELMQSRVQSGSLSSAARSEHVESEIIDGSEQENLQKMRQL
AVIPMLYDADQRIKFINMGIMDDPMKYQDRQVTNMSQEDRTFEKFMHPKN
FGLASFLERKTVACVLYYLTQAKENYKSLVRSYRRRGSKQQQQQQQQQQQ
ARSSQEEKEEKEKEKEKEKEKEKEKEKEKEKEKEDDTSGENHEKEAASKGRK
TANSCQRKRGRITRSMANHEETATPOOSSELSAMENNESSRTEEMETAKGGLL
EHRNWSAIARWVSKTSQCKNFYNYKRNQLDEILQOHLKMEKERNARRKKKT
PAAASEETAFPPAAEEMEASGASANEELAEAEASQASGNEVPRVSGPAVN

Qy	1332	YArgAlaIleProProGluArgHisSerProHisHisLeuLeuGluGlnHisIleAr	1352
Db	4122	CCGTGGCCATCCCGCGGAGCGACACAGCCCCCACCACCTCAAGAGCAGCACACACATCCG	4181
Qy	1352	gGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrIleuAr	1372
Db	4182	CGGGTCCATCACACAGGGATCCCTCGGTCTACTAGTGGAGGCACAGGAGACTACCTGCG	4241
Qy	1372	gArgGluAlaIleLeuLeuLeuArgGluGlyThrProProProProProProSerArgAs	1392
Db	4242	TCGGGAGGCCAAGCTCCTAAAGCGGGAGGGGACGCGCTCCCGCCCCCACCAGCCCTCACGGGA	4301
Qy	1392	pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLeuLeuLysProAlaHisG1	1412
Db	4302	CCTGACCGAGGCTTACAGACGCGAGGCCCTGGGCCCTCTGAAGCTGAAGCGCGCCCATGA	4361
Qy	1412	uGlyLeuValAlaThrValLysGluAlaG1ArgSerIleHisGluIleProArgGluG1	1432
Db	4362	GGGCTGTGGTCCACCGTGAAGGAGCGGGCGCGCTCCATCATGAGATCCCGCGCGAGGA	4421
Qy	1432	uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLeuGluGlySerIleTh	1452
Db	4422	GCTCGGCGACACGCGCCGAGCTGCGCCCTGGCGCGCGCGCTCAAGGAGGGCTCCATCAC	4481
Qy	1452	rGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAs	1472
Db	4482	GCAGGGCACCCCGCTCAAGTACGACACCGGGCGGTCCACACCTGGCTCCAAAAGACGA	4541
Qy	1472	pValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspVa	1492
Db	4542	CGTACCGTCCCTCATCGCAGCCCGCGCGGACGTTCCACCGCTGCACCCGCTGGATGT	4601
Qy	1492	lMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgPr	1512
Db	4602	GATGGCGGACCGCGGGACCTGGAAACGTGCTGTCTAGAGAGAGCCCTGAAGACCGCGCC	4661
Qy	1512	oGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProG1	1532
Db	4662	AGGACCGCCAGCAGCTCGGGGGCTCCATTCGCGCGCGCGCCCGCTATTGTCCTGA	4721
Qy	1532	uLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaG1	1552
Db	4722	GCTGGTGAAGCGCGGAGAGCCCTGACCTATGAGGAGCACCGGGCACCCCTTTGCGCG	4781
Qy	1552	yHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluG1	1572
Db	4782	CCACCTCCAGAGGTTCGCCGTGACCATCGGGAGCCACCGCGCGCTGACAGAGGG	4841
Qy	1572	ySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluI1	1592
Db	4842	CAGCCTTTCTGTCAGCAAGGCATCCAGGACCGGAAAGCTGACGTCGACGCTCGTGAGAT	4901
Qy	1592	eAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerProTyrG1	1612
Db	4902	CGCCAAGTCCCGGCACAGCACCGTGGCGGAGCACCCACCCCATCTCGCCCTATGA	4961
Qy	1612	uHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPh	1632
Db	4962	GCACCTGCTTCGGGGCGTGAAGTGGCGTGAGCTGTATCGAGCCACATCCCTTCGGCTT	5021
Qy	1632	eAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuPr	1652
Db	5022	CGACCCCACTCCATACCCCGCGCATCCCTCTGGAGCGAGCGCTGCCTACTACCTGCC	5081
Qy	1652	oArgHisLeuAlaProAlaProThrTyrProHisLeuTyrProProTyrLeuIleArgG1	1672
Db	5082	CCGACACCTGGCCGCCCAACCCACCTACCGGCACCTGTATCCACCTACTCATCCCGCG	5141
Qy	1672	yTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSe	1692
Db	5142	CTACCCCGACACGGCGCGCTGGAGACCGGACCATCATCATGACTACATCATCCTC	5201
Qy	1692	rGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgG1	1712

Db	5202	GCAGCAGATCCACCAACACCGCCACCGCCATGGCCCGCAGAGCTGATATGCTGAGGG	5261
Qy	1712	yLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleI1	1732
Db	5262	CCTCTCGCCCCCGAGTCTCGCTGGCATCTAACTACGCTGCGGGTCCCGAGGATCAT	5321
Qy	1732	eAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAl	1752
Db	5322	CGACCTGTCCCAAGTCCACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	5381
Qy	1752	aThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHi	1772
Db	5382	CACCGCATGGAGCCGCTTGTCTTACCTCCACCGCGCCCGCCCTTCAGCAGCGGCCA	5441
Qy	1772	sSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrSe	1792
Db	5442	CAGACCTCCCACTCTCCCGAGAGGTCCACACACTTGACANAAACCAACACACAGTC	5501
Qy	1792	rSerSerGluArgGluArgAspArgGluArgAspArgAspArgGluArgGluGly	1812
Db	5502	CTCGTCGAGCGGAGAGACCGGATCGAGAGCGGACCGGGATCGGAGCGGGA	5561
Qy	1812	sSerIleLeuThrSerThrThrValGluHisAlaProIleThrArgProGlyThrG1	1832
Db	5562	GTCCATCTCTCACCCACGACCGGTGGAGCAGCACCCATCTGGAGACCTGTGTACAGA	5621
Qy	1832	uGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly	1852
Db	5622	GCAGACGCGGAGCAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	5681
Qy	1852	aSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuG1	1872
Db	5682	CTCCACTCCCATGCCACACGACCTCGCCCATCTCCCTCGGACCCAGGATCCCTCCA	5741
Qy	1872	nGlnArgProSerValLeuHisThrGlyMetLysGlyIleIleThrAlaValGluPr	1892
Db	5742	GCAGACACCGAGTGTCTTCAACAACAGCAGCATGAAGGGTATCATCACCGCTGTGGAGCC	5801
Qy	1892	oSerLysProThrValLeuArgSerThrSerSerSerProValArgProAlaAlaTh	1912
Db	5802	CAGCAAGCCACCGTCTGAGGTCCACCTCCACCTCTCTACCGTTGCGCCAGCTGCCAC	5861
Qy	1912	rPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe	1932
Db	5862	ATTCCCACTGCCACCCACTGCCCCACTGGCGGCACTCGATGGGTCTTACCTACCT	5921
Qy	1932	uMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProAr	1952
Db	5922	CATGGAGCCGCTTGTCTGCCAAGGAGGCCCCCGGGTGGCCGCGCAGAGCGGCCCCG	5981
Qy	1952	gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl	1972
Db	5982	AGCAGACACCGGCATTCCTCTCCCAAGCCGCCACCGCCGCTCGGGCTGGAGCCCGC	6041
Qy	1972	aSerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl	1992
Db	6042	CTCCTCCCCCAGCAAGGGCTCGAGCCCCCGGCCCTTAGTGCTCTCTGTCTCTGTGCCACGC	6101
Qy	1992	aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr	2012
Db	6102	CACCATCGCCCGCACCCCTCGCAAGAACCTCGCACCTCACCGCAGCCCGCAGCCCGCC	6161
Qy	2012	oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSe	2032
Db	6162	GGCGCCACCTGCTCGGCTCGGACCGGACCGGGAAAAGACTCAAGTAAGTAAACCTTTTC	6221
Qy	2032	rIleGlnGluLeuGluArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG1	2052
Db	6222	CATCCAGGAACCTCGTCTCTGGGTATACCGCGGAGGAGCTACAGCCCGGAGG	6281
Qy	2052	yValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeuProLy	2072

Db 1954 AGCCCGAGCTGCCTCCATGAGTGAATGAGAGTTCTCGCTGCAGAGAGAAATG 2013
QY 621 GluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaAsgMet 640
Db 2014 GAACAGCCAAAGAAAGGTCTCTGGAAACACGGCCGACACTGGTTCGGCCATCGCCCGATG 2073
QY 641 ValGlySerLysThrValSerGlnCysLysAspPheTyrPheAsnTyrLysLysArgGln 660
Db 2074 GTGGGCTCCAAAGACTGTGTGCGAGTGTAAAGAACTTCTACTTCAATACAAAGAGGCGAG 2133
QY 661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetCylusGluArgAsnAlaAsg 680
Db 2134 AACCTCATGAGATCTTGCAGCAGCAACAGCTGAAGATGAGAGAGAGAGAACCGCGCG 2193
QY 681 ArgLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProValVal 700
Db 2194 AGNAGNAGAAAGAGCGCGCGCGCGCCAGCGAGGAGCTGCATTCGCGCCCGTGGTG 2253
QY 701 GluAspGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
Db 2254 GAGGATGAGGAGATGGAGGGCTCGGGCGTGACGGGAAATGAGGAGGAGATGTGTGAGGAG 2313
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
Db 2314 GCTGAA-----GCC 2322
QY 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
Db 2323 ACTGTCAACACAGCTCAGACCGAGAGCATCCCTCTCTCTCCTCAGCTGAGCGCGCCNAG 2382
QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
Db 2383 GACACAGGGCAGAATGGGCCCAAGCCGCCAGCCACCTGGGGCGCCAGCGGGCCACCCCA 2442
QY 781 GlyProProThrProPro-ArgArgThrSerArgAlaProIleGluProThrProLase 800
Db 2443 GGGCCACCCACCCACCCAGCGAGGACATCCCGGCC-CCCATGAGTCCACCCCGGCCCTC 2501
QY 800 rGluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProProPr 820
Db 2502 TGAAGCCACCTTAGCCCTTAGCCGCCCCACGACGACCCCATTTTCCCTCTTCACCTCTCC 2561
QY 820 oValValProLysGluGluLysGluGluGluThrAlaAlaAlaProProValGluGlu 840
Db 2562 TGTGGTCCCCAAG 2621
QY 840 yGluGluGlnLysProProAlaAlaGluLeuAlaValAspThrGlyLysAlaGlu 860
Db 2622 GGAGGAGCAGAACCCCGCGCTGAGGAGCTGGCAGTGGACACAGGGAGGCGCGAGGA 2681
QY 860 uProValLysSerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAl 880
Db 2682 GCCCGTCAAGAGCGAGTGCACGGAGNAGCCGAGGAGGCGCGCCNAGGCGAAGAGCGC 2741
QY 880 aGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyLys 900
Db 2742 GGAGGCGCTGAGGCCACCGCCGAGAGGCGCTCAAGGCGAGAGAGAGAGAGAGAGAGAG 2801
QY 900 rGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaTh 920
Db 2802 CGCAGGGGCCACACAGCCAAAGAGCTGGCGCGCGCCCGCCAGGACAGCGACTCCAGTGCAC 2861
QY 920 rCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPr 940
Db 2862 CTGAGTGCAGACGAGGTGATGAGGCCGAGGGGGCGGCAAGAACCGGCTGTGTCTCCC 2921
QY 940 oArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPr 960
Db 2922 AAGGCCAGCCCTCTCACCCCGAGCTGGCGAGCCCGCGGCGCAATGTCCTCACCCAGAGGCC 2981
QY 960 oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaIleProProIleGlnValThrLys 980
Db 2982 ACTGGACCTGAAGCAGCTGAAGCAGCGAGCGGCTGCATCCCGCCCATCCAGGTCAACAA 3041

QY 980 sValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPr 1000
Db 3042 AGTCCATGAGCCCGCCCGGAGGAGCGAGCTCCACCAAGCCAGCTCCCGCCAGCCACC 3101
QY 1000 oProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProAr 1020
Db 3102 GCCACCGCAAAACCTGCGAGCGGAGAGCGCCCTCAGCAGCTGGCAGCGAGCCCGG 3161
QY 1020 gGlyLysSerArgSerProAlaProProAlaAspLysGlu----- 1033
Db 3162 GGGCAAGACAGAGAGCGCGCCACCCCGCCGACAAAGAGGAGGAGAGAGCGCTGTGTCTT 3221
QY 1034 ----AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGl 1052
Db 3222 CCAGCCTTTCGAGCGCGAGCGCCAGAAAGCTGCTGGGAGCCCGCTTGTGGACTTCCGG 3281
QY 1052 yLeuProPheProValProProArgGluValLysAlaIleLysAlaSerProHisAlaProAspPr 1072
Db 3282 CCTGCCCTTTCGCGTGCCTCCCGTGGAGTGATCAAGGGCTCCCGCATGCCCGGAGCC 3341
QY 1072 oSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAl 1092
Db 3342 CTCAGCCTTCTCTAGCCTCCACTGGTCACTGCCCTGGGCTTCATGACACTGC 3401
QY 1092 aArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSerAl 1112
Db 3402 CCGGCCGCTGCTCGCGCGCCACCCACCATCTCCAAACCGCTCCCTCATCTCTCTGC 3461
QY 1112 aLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGl 1132
Db 3462 CAAGCACCCAGCGCTCTCGAGAGGCAATAGTGTCCATCTCCCAAGGAATGTGCGTCCA 3521
QY 1132 nLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLe 1152
Db 3522 GTTCCACGTCCCGTACTCAGAGCATGCCAAGGCCCGGTGGGCCCTGTCCATATGGGCT 3581
QY 1152 uProLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLe 1172
Db 3582 GCCCTGCTCCATGGACCCCAAAAGAGTGCACCTTCAGCGGAGTGAAGCAGGAGCAGCT 3641
QY 1172 uSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAl 1192
Db 3642 GTCCCAACCGGCGCCAGGCTGGGCCACCGAGAGCCTGGGGGTGCCACAGCCAGGAGGC 3701
QY 1192 aSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyI 1212
Db 3702 GTCCGTGCTGAGAGGAGCAGCTCTGGGCTCAGTTCCGGCGGAGAGCATCACCAAGGCAT 3761
QY 1212 eProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGl 1232
Db 3762 TCCAGCACACGGGTGCTCGAGACAGCGCCATCACATACCGCGGTCTCATCCACCGG 3821
QY 1232 yThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerPr 1252
Db 3822 CACGCAGCTGAGTCTCTGTACAAAGGCGCACCATCACAGGATCATCGCGAGGAGACGCC 3881
QY 1252 oSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGl 1272
Db 3882 GAGTCTCTTGGACCGCGCGCGGAGGAGACAGCTGCCCAAGGGCCAGCTCATCTACGAGG 3941
QY 1272 yLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGl 1292
Db 3942 CAAGAAGGCCACGCTCTGTCTATGAGGGTGGCATGTGTGTGACCATGTCTCCCAAGGA 4001
QY 1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrThrAs 1312
Db 4002 GGACGGCAGAAGCAGCTCAGGACCCCGCCCATGAGACGGCGCGCCCGCCCAAGCGACCTATGA 4061
QY 1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGl 1332
Db 4062 CATGATGAGGGCGCGGTGGGAGAGCCCATCTCTCAGCGAGCATCGAAGGTCTCATGGG 4121

SHSHAHQHSPISPRTODALQORPSVLNHTGMKGIITAVEPSKPTVLSTSTSSVRPA
ATFPFPAHCPGLGTLGVPTLMEPVLLKPEAPRVARPERPADTGHAFAPPARSG
LEPASPSKSESEPRPLVPVSHATARTPAKNALPHASPPDPAPPASGAPREKRT
QSKPFSFOELERLGLVHGSYSPEGVSPVSPSLTHDKGLPKHUEELDKSHLEG
ELAPKOPGVKLGEEAHLPHLPPLPESQPSPLIQTAPGVKHQRVVTLAQHISEV
ITODYTRHPHQQLSAPLPAPLXSPFCASCPVLDLRRPSPDLPLPPDHGAPARGSPHS
EGKRSPEFNKTSVLGGEDGIBPVSPPEGMTEPHRSADVPLLYRDGEOTEPSRMG
SKSPGNTSQPPAFSKLSTESANWVSKKQBEINKKUNKLTINREPEPINISQPGTEIFNM
PAITGTGLMYTRSQVQEHASTNMLEAIIRKALMGKIDOWEESPISLSANAFNLNAS
ASLPAPAMPITAAADRGSDHDLTSPGGGKAKVGRSPSRKAKPAPGLASGDRPPSVSS
VHSEGDNRRTPLTNRWEDRSPSSAGSTPPPNPLIMRLQAGVMA5PPPPGLPAGSGP
LAGPHHAWDEEPKLLCSQVETLSDSE"

ORIGIN		
Alignment Scores:		
Pred. No.:	1,26e-254	Length: 8686
Score:	12978.00	Matches: 2487
Percent Similarity:	98.57%	Conservative: 3
Best Local Similarity:	98.46%	Mismatches: 9
Query Match:	98.21%	Indels: 28
DB:	9	Gaps: 3
US-09-522-753-5 (1-2517) x AF125672 (1-8686)		
Qy	1	MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db	157	ATGTGGGCTCCACACAGCCTGTGGCACAGACGTGGAGGGCCACTGAGCCCGCTACCCG 216
Qy	21	ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
Db	217	CCCCACAGCCTTCTACCCAGTCAGATGCCCGGACGACACGGACGTCGGGCTCCTG 276
Qy	41	GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln 60
Db	277	GAGTACCAGCACCACTCCCGGAGCTATGCTCCACCTCGCCCGGCTCCATCATCCAG 336
Qy	61	ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
Db	337	CCCCAGCGCGAGGCGCTCCCTGCTGTCTGAGTTCAGCTCCGCGGAGTAACGGTCCAG 396
Qy	81	GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
Db	397	GAGCTCCACCTGGCGCCAGAGTCCCACTCATACCTGCCCGAGCTGGGGAGTCAGAGATG 456
Qy	101	GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db	457	GAGTTCATTGAAGCAAGCGCCCTCGGTAGAGTGTGCTGAGCTGCTGACCCCTCTCGGACCG 516
Qy	121	SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db	517	TCACCCCTGTGCGCCACCGGCGAGCTGGGGATCTGAAGACCTCACCAAGGACCGTAGC 576
Qy	141	LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
Db	577	CTGACGGGCAAGCTGGAACCGGTGTCTCCCGCCAGCCCGCCGACACTGACCCCTGAGCTG 636
Qy	161	GluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAsp 180
Db	637	GAGCTGGTGGCCGCCACCGCTGTCCAGGAGGAGCTGATCCAGACATCGACCGGTGGAC 696
Qy	181	ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu 200
Db	697	CGAGAGATCACCATGTTAGAGCAGCAGATCTCTAAGCTGAAGAGAAGCAGCAACAGCTG 756
Qy	201	GluGluGluAlaAlaLysProProGluProGluLysProValSerProProIleGlu 220
Db	757	GAGGAGGAGGCTGCCAAGCGCCGAGCTGAGAGCGCGGTGTACCCCGCCCACTCGAG 816
Qy	221	SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAla 240
Db	817	TGAAGACCGGAGCCTGTGTGAGATCATCTACGACGAGACCCGGAAGAGGCTGAAGCT 876
Qy	241	AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260

Db	877	GCACATCGGATTCGTGAAGGCGCTGGGGCCCGCAGGTGGAGCTGCGCTGTGTACACGAGCCC 936
Qy	261	SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
Db	937	TCCGACACCCCGCAGTATCATGAGAACATCAAAATAAACACAGGCGATCGGAGAGAGCTA 996
Qy	281	IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
Db	997	ATCTTGTACTTCAAGAGGAGGAAATCACCGCTCGAAACAAATGGGAGCAGAAAGTTCTGCCAG 1056
Qy	301	ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
Db	1057	CGCTATGACCACTCATGGAGGCGCTGGGAGAGAAAGGTGGAGCGCATCGAACAACCCC 1116
Qy	321	ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
Db	1117	CGGCGCGGGCCCAAGAGAGCAAGGTTCCGAGTACTACGAGAACAGTTCCTCGAGATC 1176
Qy	341	ArgLysGlnArgGluLeuGlnArgMetGlnSerArgValGlyGlnArgGlySerGly 360
Db	1177	CGCAAGCAGCGCGAGCTGCAGGAGCGCATGCAG---AGGGTGGGCGCAGCGGCGCAGTGGG 1233
Qy	361	LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
Db	1234	CTGTCCATGTCGCCCGCCGCGCAGCAGCAGGAGTGTGAGAGATCATCGATGGCTCTCA 1293
Qy	381	GluGlnGluAsnLeuLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyr 400
Db	1294	GAGCAGGAGAACTCGAGAGCAGATGCGCCAGCTGCGCGTGTATCCCGCCCATGCTGTAC 1353
Qy	401	AspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet 420
Db	1354	GAGCTGACCAAGCGCATCAAGTTCAATCAATCAAGCGGCTTATGTCGCCACCCCATG 1413
Qy	421	LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPhe 440
Db	1414	AAGGTGTACAAAGACCGCCAGGTCATGAACATGTGGAGTGAGCAGGAGAGGAGACCTTC 1473
Qy	441	ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
Db	1474	CGGAGAGAGTTCATGAGCATCCCAAGAACTTTGGCTGATCGCATTCCTCTGGAGAGG 1533
Qy	461	LysThrValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480
Db	1534	AACACAGTGGCTGAGTGGCTCTCTATTACTACTGACTGACTAAGAAGAATCAGAACTATAAG 1593
Qy	481	SerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
Db	1594	AGCTGTGAGACGGAGCTATCGGCGCGCGGAGAGCCAGAGCCAGCAGCAACACGACAGCAG 1653
Qy	501	GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db	1654	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1713
Qy	521	AspGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsn 540
Db	1714	GATCAGAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1773
Qy	541	AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu 560
Db	1774	GACAGGAAGACCTCTCTCAGGAGAGAGCAGACACCTCAGGGGAGGAGCAGCAGCAGCAG 1833
Qy	561	LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
Db	1834	AAGGAGGCTGTGGCTTCAAGGCGCAAACTCCCAACAGCCAGGAGGAGGAGGAGGAGGAGG 1893
Qy	581	ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db	1894	CGCATCCCCGCTCAATGGCTAATGAGGCCAACAGCGAGGAGGCCATCACCCCCCAGCAG 1953
Qy	601	SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMet 620


```
Qy 2161 AlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgProPro 2180
Db 6482 GCGCCCTCTACTCTCTCCCTGGGCGAGCTGCGCTGCTGAGCTCCGCCGCCACCC 6541
Qy 2181 SerAspLeuTyrLeuProProAspHisGlyAlaProAlaArgGlySerProHisSer 2200
Db 6542 AGTGACCTCTACTCTCCCGCCCGGACCATGGTGCGCCCGCGCTGCTCCCGACAGC 6601
Qy 2201 GluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAsp 2220
Db 6602 GAAGGGGCAAGAGTCTCCAGAGCCAAACAAGCGTCTGCTGGTGCTGGAGGAC 6661
Qy 2221 GlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAla 2240
Db 6662 GGTATTGAACCTGTGCTCCCGGAGGATGATGCGGAGCCAGGAGCTCCCGAGTGCT 6721
Qy 2241 ValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLys 2260
Db 6722 GTGTACCCGCTGCTGTACCGGATGGGGAACAGAGCGGACCGAGGATGGGCTCAAG 6781
Qy 2261 SerProGlyAsnThrSerGlnProProAlaPheSerLysLeuThrGluSerAsnSer 2280
Db 6782 TCTCCAGCAACACCGAGCGCGAGCTCTTTCAGCAAGCTGACGAGCAACTCC 6841
Qy 2281 AlaMetValLysSerLysGlnGluLeuAsnLysLysLeuAsnThrHisAsnArgAsn 2300
Db 6842 GCCATGGTCAAGTCCAGAGCAAGAGATCAACAAGAGTGAACACCCCAACCGGAAT 6901
Qy 2301 GluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThr 2320
Db 6902 GAGCCTGAATACATATACAGGAGCTGGGACGAGATCTTCAATATGCGCGCATCAC 6961
Qy 2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMet 2340
Db 6962 GGAACAGCGCTTATGACCTATAGAGCGCGAGCGGTGAGCAACATGCAACCATG 7021
Qy 2341 GlyLeuGluAlaIleAlaArgLysAlaLeuMetGlyLysTyrAspGlnThrGluGluSer 2360
Db 7022 GGGCTGGAGGCCATATATTAGAAAGGCATCATGGTAAATATGACCAAGTGGAGAGTCC 7081
Qy 2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla 2380
Db 7082 CCGCCGCTCAGCGCCAAATGCTTTTAAACCTCTGAATGCCAGTCCAGCTGCCCGCTGCT 7141
Qy 2381 MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly 2400
Db 7142 ATGCCCATACCGCTGCTGACGCGAGTGACCAACACTCACCCTGCCAGTGGCGGC 7201
Qy 2401 GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGly 2420
Db 7202 GGAAGGCCAAGGTCTCTGGCAGACCCAGCAGCGGAAAGCCAAAGTCCCGCGCGCGGC 7261
Qy 2421 LeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCysAsn 2440
Db 7262 CTGGCATCTGGGACCGCGCCACCTCTGCTCTCTCAGTGACCTCGGAGGAGACTGCAAC 7321
Qy 2441 ArgArgThrProLeuThrAsnArgValTyrGluAspArgProSerSerAlaGlySerThr 2460
Db 7322 CGCGGACCGCGCTCACCAACCGGTGTGGGAGGACAGGCGCTCGTCCGAGGTCCACG 7381
Qy 2461 ProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro 2480
Db 7382 CCATTCCCTTACAAACCCCTGTATCATCGGCTGCGAGCGGCTGTATGGCTTCCCGACCC 7441
Qy 2481 ProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaTyrAspGlu 2500
Db 7442 CCACCGGCGCTCCCGGGGAGCGGGCGCCCTGCTGCGCGCCCGCCACCAACCGCTGGAGAG 7501
Qy 2501 GluProLysProLeuLeuCysSerGlnTyrGlnThrLeuSerAspSerGlu 2517
Db 7502 GAGCCCAAGCACTGCTCTGCTCGAGTACGAGACACTCTCCGACAGCGAG 7552
```

```
AF125672
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
```

```
AF125672      8686 bp      mRNA      linear      PRI 04-APR-1999
Homo sapiens silencing mediator of retinoic acid and thyroid
hormone receptor extended isoform (SMRTE) mRNA, complete cds.
AF125672.1   GI:4559297
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 8686)
Park,E.J., Schroen,D.J., Yang,M., Li,H., Li,L. and Chen,J.D.
SMRTE, a silencing mediator for retinoid and thyroid hormone
receptors-extended isoform that is more related to the nuclear
receptor corepressor
Proc. Natl. Acad. Sci. U.S.A. 96 (7), 3519-3524 (1999)
99139215
10097068
2 (bases 1 to 8686)
Direct Submission
Chen,J.D.
Submitted (03-FEB-1999) Pharmacology and Molecular Toxicology,
University of Massachusetts Medical School, 55 Lake Avenue North,
Worcester, MA 01655, USA
Location/Qualifiers
1..8686
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..8686
/gene="SMRTE"
157..7680
/gene="SMRTE"
/function="transcriptional corepressor"
/notes="SMRTE"
/codon_start=1
/product="silencing mediator of retinoic acid and thyroid
hormone receptor extended isoform"
/protein_id="AAD2973.1"
/db_xref="GI:4559298"
/translation="MSGSTQPVAAQTWRATEPRYPHSLSYPVQIARTHTDVGLEVQH
HSDYASHUSPGSIIQPKRRRLSESEFQPGNRSQELHRSFHSYLPGLKSENEF
IESKPRLELLPRLPLATQPAQSEDLTKDRSLTKLEPVSPVPHDDEL
ELVPRLSKEELIQNDREITWVEQIISKLKKQQLQLEBEAAKPPPEKVPSP
IESKRSIVIIYDNRKKAARHILEGLQVQLPLYNQPSDTRQYHENTKINAM
RKLLILFKRNHARKQWQKFCQYDOLMEAWKVERIENPRRAKESKVREYE
KQFPKIRKQRELQRMQVQSGSLSPARSEHEVSEIILQSEQENLEKQMRQLA
VIPLMIYDADQRIKPIINNGLMADPMKVYKDKQVNMWSEQEKETFRFKFWQPKNF
GLIASFLEKTVACVLYYLTKNENYKSLVRSYRRKSGKQQQQQQQQQQQQQ
QMPRSSQDEKDEKEAEKEBEKVEVNDKEDLLKEKTDGTSDEDEKAEVASK
GRKTANSQGRKGRITRSMANEANSEEAITPOQSAELASMLNESRMTSEMETAKK
GILHCRNWSATARWGSKTVCNKFYNYKKRONLDEILOHKLKMEKERNARKK
KAPAAASEEAPPPVEDEMEASGVTGNEEMVEEAEATVNNSDTSISPSPTTEA
ADKTGNGPQATGADGPPGPPPTPPEDIPAPTEPTEATLAPTPPPAPPPSP
SPFPVYVEKEEETAAAPPEVEGEEKPPAAELAVDTGKAEVPKSECTSEAESEGP
AKGDAEAEATAERALKAEKGGSGRATTAKSGAPQSDSSATCSADEVDEAEGG
DKNRLSPRLSLTPTDPRANAPKQPLDLKQKORAAAIPPIQVTKVHEPREDA
PTKPPAPPPQNLPQESDAPQPGSSPRGSRSPAPPADKAEKVPFPFAAQAQ
KLPDPPTWTSGLPPVPREVIKASPHAPDSAFSPAPGHPGLGLDHTARPLYPR
PTIINPPPLISSAKHPSVLERIQIAGISQMSVOLHVPYSEHAKAPVGPVTMGLPLPM
DPKLPAGSGVQEQSLPRGQAGPPESLGPVTAQASVLRGTLGSDGSLPKGHIYEG
TRVPSDSATITRGSITHTGPDADLYKGTITRIIGSDPSRLDRGRDLSLPGHIYEG
KGHVLSYBGMSVTQCKEDGRSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGL
MGRATYPPEHSPHLKEOHIRGSIQTQIPRSYVAAQEDYLREAKLKREGTPPPPP
PSRDLTEAYKTOALGPLKLPKPAHEGLVATVKEAGRSIHEIPREELHTPELPLAPRL
KESGSIQTGPLKIDTASTGTGSKHDDVRLSGSPGRTFPFVPLDPMADALAGERAT
ESLSKRPQTASSGSGIARGAPVIVPELGRQPSGLTYFDHGAPGAPVPMGSLPVTM
REPTPRLQEGSLSSKASQDRKLTSTPREIAKSPHSTVPEHHPHTISPYHLRGVSG
VOLYRSHIPLAFDPTSIPRGIPLDAAAAYLPHLAPNPTPHLYPYLIRGVPDFAA
LENROTIINDYITSOQMHNTATAMAQRAQMDLRLGSLPRESSLALANYAAGPRGII
VHLPLVLPPTGCTPATMDRLAYLPTAPQFPSSRRHSSPLSPFGGTHLTIKPTTSSS
ERDRDRDRDRDREREKSILTSTTTTVEHAPIWRPTEQSSGSSGSSGSSGSSRPA
```


QY 701 GluAspGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGlu 720
DB 2102 GAGGATGAGGAGTGGAGCGTGGCGTGGCGGAAATGAGGAGGAGATGGTGGAGGAG 2161
QY 721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla 740
DB 2162 GCTGAAGCCCTTATGCTCTGGGATGAGTGGCCAGAGGGGAATGCACTGGCGCCAGCC 2221
QY 741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys 760
DB 2222 ACTGTCAACACAGCTCAGACACGAGAGATCCCTCTCTCCTCCTCAGCTGAGGCGGCCAG 2281
QY 761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro 780
DB 2282 GACACAGGCGAGAAATGGGCGCAAGCCCGCCAGCCAGCCCTGGGCGCCAGCGGCCACCCCA 2341
QY 781 GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSer 800
DB 2342 GGCCCAACCCACCCACACGAGGACATCCCGGCGCCCAATTGAGCCCAACCCCGGCTCT 2401
QY 801 GluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProProPro 820
DB 2402 GAAGCCACCGAGCCCTAGCCCCCCACACGACACCCCTCGCCCTCTGACCTCTCTCT 2461
QY 821 ValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGluGly 840
DB 2462 GTGGTCCCAAGGAGGAGGAGGAGGAGACCCGACAGCGGCCCGCCAGTGGAGGAGGG 2521
QY 841 GluGluGlnLysProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGlu 860
DB 2522 GAGGAGCAGAAGCCCGCGGCTGAGGAGCTGGCAGTGAGACACAGGAAAGCGCGAGGAG 2581
QY 861 ProValLysSerGluCysThrGluGluAlaGluGlyProAlaLysGlyLysAspAla 880
DB 2582 CCGCTCAAGACGAGTGACGAGGAGGAGCCGAGAGGGGCGGCCCAAGGCGCAGGCGG 2641
QY 881 GluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlySer 900
DB 2642 GAGGCGCTGAGGCCACCGCGGAGGCGGCGCTCAAGCAGAGAAGGAGGCGCGGAGC 2701
QY 901 GlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThr 920
DB 2702 GGAGGCGCCACCTGCTCAAGAGCTGGGCGCGCCCGCCAGACAGCACTCCAGTGTCTACC 2761
QY 921 CysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro 940
DB 2762 TGCACTGCACACGAGTGGATGAGCGCGAGCGGCGGCGACAAACCGGCTGTCTCCCA 2821
QY 941 ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPro 960
DB 2822 AGGCCAGCCCTCTCACCCCGACTGGCGACCCCGGCGCAATGCCCTCACCCAGAGCCA 2881
QY 961 LeuAspLeuLysGlnLeuLysGlnArgAlaAlaIleProProIleGlnValThrLys 980
DB 2882 CTGGACCTGAAGCAGCTGAAGCAGCGAGCGGCTGCCATCCCGCCCATCCAGGTCAACAA 2941
QY 981 ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro 1000
DB 2942 GTCAATGAGCCCGCGGAGGAGCGAGCTCCACCAAGCAGGCTCCCGCCAGCCCGCCAG 3001
QY 1001 ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg 1020
DB 3002 CCACCGCAAAACCTGCAGCGCGGAGAGCGAGCGCCCTCAGCAGCTGGCAGCAGCCCGCG 3061
QY 1021 GlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln 1040
DB 3062 GGAAGAGCAGGAGCGCGGACCCCGCGCGCAAGAGAGGCTTCGCGAGCGCGAGGCCAG 3121
QY 1041 LysLeuProGlyAspProProCysTtpThrSerGlyLeuProPheProValProProArg 1060
DB 3122 AAGCTGCTGGGAGACCCCTTGTGTGACCTTCGGCGCTGCCCTTCCCGGCTCCCGCGT 3181
QY 1061 GluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyAlaProPro 1080

DB 3182 GAGGTGATCAAGGCGCTCCCGCATCCCGGAGCCCTCAGCCCTTCTCTACGCTCCACCT 3241
QY 1081 GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPro 1100
DB 3242 GATCACCCACTGCCCTTGGGCTTCCATGACACTGCCGCGCGCTCTCGCGGCCACCC 3301
QY 1101 ThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArg 1120
DB 3302 ACCATCTCAACCCCGCTCCCTCTCTCTCTGCAAGCACCCAGCGCTCTCGAGAGG 3361
QY 1121 GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTySerGluHis 1140
DB 3362 CAATAGTGTGCATCTCCCAAGGAATGTCGGTCCAGCTCCAGTCCCGTACTCAGAGCAT 3421
QY 1141 AlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys 1160
DB 3422 GCACAGGCCCGGTGGGCTGTACCATGGGGCTGCCCTGCCCATGAGCCCCCAAAAG 3481
QY 1161 LeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro 1180
DB 3482 CTGGCACCCCTTCAGCGGAGTGAAGCAGGAGCAGCTGTCCCAAGCGGCGCGCTGGGCCA 3541
QY 1181 ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu 1200
DB 3542 CCGGAGAGCTGGGGTGGCCACAGCCCAAGGAGGCTCGCTGCTGAGAGGAGCAGCTCTG 3601
QY 1201 GlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAsp 1220
DB 3602 GGTCTAGTTCGGGCGGAAGCATCACCAAGGATTCACAGCACCGGTGCCCTCGGAC 3661
QY 1221 SerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspValLeuTyLys 1240
DB 3662 AGCGCCATCACATACCGGCTCCATCACCCAGCAGCAGCTCAGCTCTGTACAAG 3721
QY 1241 GlyThrIleThrArgIleIleGlyLeuAspSerProSerArgLeuAspArgGlyArgGlu 1260
DB 3722 GGCACCATCACAGGATCATCGCGAGGACAGCCCGAGTCTGGACCGCGCGCGGAG 3781
QY 1261 AspSerLeuProLysGlyHisValIleTyGlyGlyLysGlyHisValLeuSerTyLys 1280
DB 3782 GACAGCTGCCCAAGGCCACGTCATCTACGAAGGCAAGAGGCCACGCTCTGTCTAT 3841
QY 1281 GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerGlyPro 1300
DB 3842 GAGGTGGCATGTCTGTGACCCAGTCTCCAAGGAGGACGCGAGAAGCAGCTCAGGACCC 3901
QY 1301 ProHisGluThrAlaAlaProLysArgThrTyArgMetMetGluGlyArgValGlyArg 1320
DB 3902 CCCCATGAGAGCGCGCCCAAGCGCCTATGACATGATGAGGCGCGCGTGGGCGAG 3961
QY 1321 AlaIleSerSerAlaSerIleGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
DB 3962 GCATCTCTCAGCCAGCATCGAGGTCTATGGGCGCTGCCATCCCGCGGAGCGACAC 4021
QY 1341 SerProHisLeuLysGluGlnHisIleArgGlySerIleThrGlnGlyIlePro 1360
DB 4022 AGCCCCCACCCCTCAAGAGCAGCAGCACCATCCCGGGTCCATCACAAAGGATCCCT 4081
QY 1361 ArgSerTyValGluAlaGlnGluAspTyLeuArgArgGluAlaLysLeuLysArg 1380
DB 4082 CGGTCTAGCTGGAGGACAGGAGGACTACCTGCTCGGAGGCCAAGCTCTCTAAAGCGG 4141
QY 1381 GluGlyThrProProProProProSerArgAspLeuThrGluAlaTyLysThrGln 1400
DB 4142 GAGGCGACGCTCCCGCCCCCACCGCTCAGCGGACCTCAGCGAGGCTTACAGAGCGAG 4201
QY 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGlu 1420
DB 4202 GCCCTGGGCCCCCTGAAGCTGAAGCGGCGCCATGAGGGCTGGTGGCCACGGTGAAGGAG 4261
QY 1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440

Alignment Scores:		
Pred. No.:	2e-259	Length: 8561
Score:	13215.00	Matches: 2517
Percent Similarity:	100.00%	Conservative: 0

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 15, 2004, 22:23:25 ; Search time 20448 Seconds
(without alignments)
5335.212 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQWTRATEPRP.....WDEPKPLCSQVETILSDSE 2517

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO.spool/US09522753/runat_15042004_143737_17428/app_query.fasta_1.2695
-DB=GenEmbl -OPMT=fastap -SUFFIX=xge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=100
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=100 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09522753@cgn 1_14601@runat_15042004_143737_17428 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.vi.*
14: gb.vl.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.oz.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Query		Length	DB	ID	Description
			Match	%				
C	1	13215	100.0	8561	9	AF113003	AF113003 Homo sapi	
	2	12978	98.2	8686	9	AF1125672	AF1125672 Homo sapi	
	3	11067	83.7	8388	10	AF113001	AF113001 Mus muscu	
	4	10832	82.0	8544	10	AF125671	AF125671 Mus muscu	
	5	9894	74.9	7465	10	AF113002	AF113002 Mus muscu	
	6	7885	59.7	5989	9	HSU37146	U37146 Human silen	
	7	4612	34.9	2930	9	S83390	S83390 T3 receptor	
	8	4194	31.7	7949	9	AB028970	AB028970 Homo sapi	
	9	4187	31.7	7780	10	MMU35312	U35312 Mus musculu	
	10	4147	31.4	7940	6	AX578069	AX578069 Sequence	
C	11	4147	31.4	7940	9	AF044209	AF044209 Homo sapi	
	12	4043	30.6	8018	9	AF087856	AF087856 Homo sapi	
	13	4009	5	30.3	8959	5	AF495886	AF495886 Xenopus 1
	14	3736	5	28.3	194840	2	AC139377	AC139377 Mus muscu
	15	3479	5	26.3	254449	2	AC097560	AC097560 Rattus no
	16	3445	26.1	218129	2	AC136560	AC136560 Rattus no	
	17	3344	25.3	219339	2	AC121005	AC121005 Rattus no	
	18	3178	24.0	2842	9	BC004326	BC004326 Homo sapi	
	19	3012	22.8	2964	10	BC047524	BC047524 Mus muscu	
	20	2891	5	21.9	205283	9	AF073916	AF073916 Homo sapi
C	21	2187	16.5	3120	9	AF303586	AF303586 Homo sapi	
	22	2155	16.3	1917	5	BC054296	BC054296 Xenopus 1	
	23	2091	5	15.8	6328	10	AB093281	AB093281 Mus muscu
	24	1900	14.4	6541	9	AB019524	AB019524 Homo sapi	
	25	1839	13.9	4285	10	MUSRI13G	L78294 Mus musculu	
	26	1687	12.8	1891	9	BC050594	BC050594 Homo sapi	
	27	1663	5	12.6	1850	9	BC058511	BC058511 Homo sapi
	28	1633	12.4	1741	9	BC056862	BC056862 Homo sapi	
	29	1628	5	12.3	3025	5	BC049302	BC049302 Danio rer
	30	1396	10.6	161970	2	AC027706	AC027706 Homo sapi	
C	31	1391	10.5	2914	9	AK127788	AK127788 Homo sapi	
	32	1272	9.6	3575	9	HSM802410	AL137641 Homo sapi	
	33	1172	5	8.9	3997	9	AF303585	AF303585 Homo sapi
	34	1162	8.8	133947	5	AL590153	AL590153 Zebrafish	
	35	1113	8.4	650	6	AF175223	AF175223 Sequence	
	36	1099	8.3	11296	3	AF175223	AF175223 Drosophil	
	37	1076	8.1	872	6	BD21548	BD21548 Human gen	
	38	960	7.3	555	6	AX677866	AX677866 Sequence	
	39	932	7.1	560	9	HSU80750	U80750 Homo sapien	
	40	932	7.1	560	9	HSU80761	U80761 Homo sapien	
C	41	880	6.7	718	6	AX753058	AX753058 Sequence	
	42	845	6.4	710	6	BD021376	BD021376 Novel gen	
	43	845	6.4	710	6	BD101314	BD101314 Novel gen	
	44	774	5	5.9	673	6	AX677865	AX677865 Sequence
	45	763	5	5.8	135351	9	AC068279	AC068279 Homo sapi
	46	760	5	5.8	150769	2	AC130376	AC130376 Homo sapi
	47	750	5.7	1793	9	AK057740	AK057740 Homo sapi	
	48	746	5	5.6	520	6	AX396270	AX396270 Sequence
	49	743	5.6	40871	2	AC020019	AC020019 Drosophil	
	50	743	5.6	160440	3	AC023741	AC023741 Drosophil	

QY 2072 LysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnPro 2091
DB 1563 AAGCACTGGAGAGCTCGACAAGAGCCACCTGGAGGGGAGCTGGGCCCAAGACGCCA 1622
QY 2092 GlyProValLysLeuGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGlu 2111
DB 1623 GGCCCCGTGAAGCTTGGCGGGAGGGCCGCCACCTCCACACCTGGCGCGCTGCCTGAG 1682
QY 2112 SerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg 2131
DB 1683 AGCCAGCCCTCGTCCAGCCCGCTGCTCCAGACGCCGCCAGGGGTCAAGGTCAACGCG 1742
QY 2132 ValValThrLeuAlaGlnHisLeuSerGluValIleThrGlnAspTyrThrArgHisHis 2151
DB 1743 GTGGTCACTGGCCAGCAGCATCAGTGGGTCAATCACACAGGACTACACCCGCGCACAC 1802
QY 2152 ProGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCys 2171
DB 1803 CCACAGAGCTCAGCGACCCCTGGCCGCCGCCCTCTACTCTTCCCTGGGGCCAGCTGC 1862
QY 2172 ProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProProAspHisGly 2191
DB 1863 CCGCTCTGAGCTTCCGCCGCCACCCAGTGAATCTTACCTCCGCCGCCGAGACCATGGT 1922
QY 2192 AlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLys 2211
DB 1923 GCCCGCGCGCTGGCTCCCCCAGCAGGAGGGGCAAGAGGTCTCCAGAGCCAAACAAG 1982
QY 2212 ThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProProGluGlyMet 2231
DB 1983 ACCTCGCTCTGGGTGGTGGAGGAGGATTAATGAACCTGTGTCGCCACCGAGGGCGATG 2042
QY 2232 ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln 2251
DB 2043 ACCGAGCCAGGCACTCCCGAGTGTGTGTGTACCCGCTGTGTACCGGGATGGGGACAG 2102
QY 2252 ThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhe 2271
DB 2103 ACCGAGCCAGCAGGAGTGGCTCCAGTCTCCAGGCAACACAGCCAGCCGCCAGCTTC 2162
QY 2272 PheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGlnIleAsn 2291
DB 2163 TTGAGCAAGCTGACCGAGAGCACTCCGCCATGTCAGTCCAGGAGCAAGAGATCAAC 2222
QY 2292 LysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThr 2311
DB 2223 AAGAGCTGAACCAACCCAGGATGAGCTGAATACATATACAGCCAGCTGGGAGC 2282
QY 2312 GluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAla 2331
DB 2283 GAGATCTTCAATATGCCCGCCATCACCAGGAACAGGCTTATGACCTATAGAGCCAGGCG 2342
QY 2332 ValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMet 2351
DB 2343 GTGAGGAACATGCCAGCAACATGGGCTGGAGGCCATAATTAGAAAGGCACCTCATG 2402
QY 2352 GlyLysTyrAspGlnTyrGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeu 2371
DB 2402 ----- 2402
QY 2372 AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAsp 2391
DB 2402 ----- 2402
QY 2392 HisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSer 2411
DB 2403 -----GGTGGCGCGGGAAGGCCAAGGTCTCTGGCAGACCCAGCAGC 2444
QY 2412 ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSer 2431
DB 2445 CGAAAGCCAAAGTCCCCGGCCCCGGGCTGGCATCTGGGACCGGCCACCTCTGTCTCC 2504
QY 2432 SerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTyrGlu 2451

DB 2505 TCAGTGCATCGAGGAGGAGACTGCAACCCGCGACCCGCTACCAACCGCGTGTGGAG 2564
QY 2452 AspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeu 2471
DB 2565 GACAGGCCCTCGTCCGAGGTTCCAGCCATCCCTTACACCCCTGATCATGCGGCTG 2624
QY 2472 GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu 2491
DB 2625 CAGCGCGGTGTATGCTTCCCCACCCACCGCGCTCCCCCGGGCAGCGGCGCCCTC 2684
QY 2492 AlaGlyProHisHisAlaTrpAspGluProLysProLeuLeuCysSerGlnTyrGlu 2511
DB 2685 GCTGGCGCCACACACCCCTGGGACGAGGAGCCCAAGCCACTGCTCTGCTGCAGTACGAG 2744
QY 2512 ThrLeuSerAspSerGlu 2517
DB 2745 AACTCTCCGACAGCAG 2762

RESULT 8
AB028970
LOCUS Homo sapiens mRNA for KIAA1047 protein, partial cds.
DEFINITION AB028970
ACCESSION AB028970.2 GI:29421187
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)

REFERENCE 1
AUTHORS Kikuno,R., Nagase,T., Ishikawa,K., Hirose,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
JOURNAL DNA Res. 6 (3), 197-205 (1999)
MEDLINE 99397452
PUBMED 10470851

REFERENCE 2 (bases 1 to 7949)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdmainfo@kazusa.or.jp. Tel: +81-438-52-3913, Fax: +81-438-52-3914)
COMMENT On Mar 31, 2003 this sequence version replaced gi:5689430.

FEATURES
Location/Qualifiers
1..7949
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="17"
/clone="hh0122181"
/tissue_type="brain"
/notes="vector: pBC SK plus"
1..7949
/gene="KIAA1047"
<12..7058
/notes="KIAA1047"
/notes="Start codon is not identified."
This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1 - 1691 was derived from hh01221 and 1692 - 7949 was derived from hg02672).
/codon_start=1
/product="KIAA1047 protein"
/protein_id="BAA82999.2"
/db_xref="GI:29421188"
/translation="SHSLVRTLIMSSGYPNPGAFSTEQSRYPHPSVQYTFPNTR HQGEFVAVPYRSHLVSQASQLQQQQQLRRRLSEFHPGSDRPOERTSYEP FHPGSPVDHDSLESKPRLEQVSDSHFQRVSAVPLVPLPLEGLRASADAKDPAF GKGHEAPSPISGQPCGDDQNASPSKLSKEELIQSDMRDVRDREIAKVEQILLKKKKQO

QLEEEAAKPEPEKPVSPPEVEQKHSIVQIYIDENRKAEEAHKIFEGLGKVELPL
YNOPTKVVHENIKTNQVMKKLILFFKRNNHARKOREOKICORVDQJLMEAWKKEVD
RIENRRKRAKSEKTEYKQPEIRKOREQORFORVORGAGCATCCAGCATTCGGA 494
EIIDGLSEENNEKQKQSVIPPMFDEAQRVKFINNGLMEDPMKYKQKQPMV
WTDHEKIFADKQIPKQGLFIASTLERKSPDCVLYIYLTIKNENIKALVRNKGK
RGRNQOIAPEEKVEKBEKDEKKEKDEKDEKDEKDEKDEKDEKDEKDEKDEK
GTABETEERQATPRKGTANSGRKGRITRSMTEAAAAAATAAATTEPPPLPP
PPEPISTPEVTSRWTEEMEVAKGLVEHGRNAAIAKMGVTKSAQCKQPFVNRK
RHNLDLQHQKOKTKREERDVOCESVASTVGAODEDIEAENRENEDESGA
ENKSDTESAPSPVEAVKPSDESPENATSRGNTEPAVELEPTTEAPTSTSLAVPS
TKPAEDSVETQVNDISIAETAQMDVDQOQESHAERGSVCDPPPATKADSVDEVRVP
ENHASKVEGNTKRDLDKASEKVEPRDVLVVAQINQAPROPEQSDNDSSATCSADE
DVNDEPERQMFMDKSLNPTGSLVSLKPNPLDLPOLQHRAAVIPPVMSCTP
CNIPITPGSGVALYOHKIKAMESALLBEORQORQIDLECRSSTPCGTSKSPNRE
WEVLQPAHPQVITNLPEGVRLPTTRPPRPLIPSSKTTVAESKPSFTMGGSISQGT
PGTYLTSHNOASVTOETPSPVSGISLGLPROQESAKSATLPVKEEESPRSQNSQP
EGLLVRAQEGHVGRGAGALQESITRGTPTSKISVESIPSLSGSITQGTPALPQTGI
PTBALVKSISREMIEDSPKGRREEAASKGVHYEGKSHGLSYDNIKNAREGTRSP
RTAHEISLRKSYSEVGNIKQMSRESPVSALEGLICRALPRGSPHDLKERTVLS
GSMQGTFRATTSFEDGLKYPKQIKRESPPPIRAFEAGITKGPYDGTITIKEMGRSI
HEIPRODILTOBSRKTPEVVQSTRPIIEGSIQGTPIKFDNNSGQSAIKHNKSLITG
PSKLSRGMPLLEIYENIKVVERKVEDYKAGETVSRHTSVVSSGSPVLSLTHAP
KAQLSGIIIDTSARTIPVSYQNTMRSGSPMMRTSDVITSSNKNHRSKSLTPTQ
RESIPAKSPVGVDPVVSHPSPFDPHHRGTAGEVYRSHLPTLHPAMPFHRALDPA
AYLFQRLSPTPGPSQOYLAMENTROTILANDYITSQMVNLRDPAVAGLSPREQP
LGLYPATRGIIIDLTNMPPTILVHPGCTSTPMDRITVPGTOIFPPRPYNASMS
PGHPHTLAAASAEIERERERERERERERERERERERERERERERERERERER
SPSVRTQETMLQORPSVFOGTNGTSVITPLDPTAQLRIMPLPAGGSPISQGLPASRYN
TAADALAAVAAASAPQMDVSKSEISSHRYETPDSALEVISPPASPPAPPEKQTY
QPEVVAQNAENDPTROYEGLPHRYPOQSPPOQLPPSQOAGQVPRTHLIT
LADHICQIITQDFARNQVSQTPQPTSTFQNSPALSVPVTRTNSRYSPEQAO
SVHQRPGSRVSPENLVSRGSRPGKSPERSHSEPEYEPISPPPOVPVVEKODSLL
LLSORGAEPABOQRNDARSIGSISLPSFTKLENTSPMVKSKQETFRKLNSSGGSDS
DMAAOPGTEIFNLPAVTTSGSVSRGSPADPNLGLIEDIRKALNSPDKVEDH
GVMSVPMGVPGTANTSVVTSGETRREGDPSPHSGVCKPKLISKNSRKSERSIP
GGYLGTERPSSVSHSEGDYHROTPGWAWEDRPSSTQTFPPNPLTMRSLSTPP
TPIACAPSVAQNQAAPHQONRIWEREPAPLLSAQYETLSDD"

ORIGIN

Alignment Scores:			
Pred. No.:	5,63e-77	Length:	7949
Score:	4194.00	Matches:	1097
Percent Similarity:	53.92%	Conservative:	340
Best Local Similarity:	41.16%	Mismatches:	720
Query Match:	31.74%	Indels:	508
DB:	9	Gaps:	108
US-09-522-753-5 (1-2517) x AB028970 (1-7949)			
QY	16	GlupProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr	35
DB	93	CAAAAGTCGTATCTCTCTCACTCTGTCAGCATATACATTTCCCAACACCCGCCACAGCAG	152
QY	36	AspValGlyLeuLeuGluTyrGln-----HisHisSerArgAspTyrAlaSerHisLeu	53
DB	153	GAGTTCGCGAGTCCCTGATTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	212
QY	54	SerProGlySerIleIleGlnProGlnArgArgArgProSerLeuSerLeuSerGluPheGln	73
DB	213	TTCCAGCAACAGCAGCAGCAGCAGCAGTTCGAAAGCGGACCTTCTCTGCTTCAGAAATTCAC	272
QY	74	ProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPro	93
DB	273	CCAGGTTCAGCAGGCTCAAGAA-----AGGAGAACTAGTTATGAACCGTTTCATCCA	326
QY	94	GluLeuGlySerGluMetGluPheIleGluSerIysArgProArgLeuGluLeuLeu	113
DB	327	GGCCCATCCCCAGTGGATCATGATTCATCGAATCGAAGCGACCGTCTGGAAACAGGTT	386
QY	114	ProAsp-----ProLeuLeuArgProSerPro	122
DB	387	TCTGATTCTCATTTTCAGCGTGCAGTGTGCGGGTTTGTTCCTTTAGTGACCCGCTGCCA	446
QY	123	LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrIysAspArgSerLeuThr	142
DB	447	-----GAAGGGCTGAGGGCT---TCTGCAGATGCTTAAGAGGATCCAGCATTCGGA	494
QY	143	GlyIysLeuGlu---ProValSerProProSerProProHisThrAspProGluLeuGlu	161
DB	495	GGCAAAACATGAAGCTCCATCTCTCCAAATTCGGGGCAACCATGTGTGAGATGATCAAAAT	554
QY	162	LeuValProProArgLeuSerIysGluLeuLeuIleGlnAsnMetAspArgValAspArg	181
DB	555	GCTTCACTCCCTCAAACTCTCAAGGAAGAGTAAATACAGAGTATGGATCGTGTAGATCGA	614
QY	182	GluIleThrMetValGluGlnGlnIleSerIysLeuIysIysGlnGlnGlnLeuGlu	201
DB	615	GAATTTGCAAAAGTAGAACAGCAGATCTCTTAACTGAAAAAGAAACAACAACAGCTTGA	674
QY	202	GluGluAlaAlaIysProGluProGluIysProValSerProProProIleGluSer	221
DB	675	GAAGAGCAGCTAAACCTCTGAGCGTGAAGCCGCTGCTCCCTCTCTCTGTGGAGCAG	734
QY	222	LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgIysIysAlaGluAlaAla	241
DB	735	AAACCCGAGTATTGTCCAAATTTATTTATGATGAGAAATCGGAAGAAAGCAGAGAAGCT	794
QY	242	HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer	261
DB	795	CATAAAATTTTGAAGGTCTTGCCCAAAAGTTGAATCCCATCTGTATAACAGCCATCA	854
QY	262	AspThrArgGlnTyrHisGluAsnIleIysIleAsnGlnAlaMetArgIysIysLeuIle	281
DB	855	GATACCAAGCTGTACCATGAGAACATCAAGACAAACAGGATGATGAGGAAAAAACTCAT	914
QY	282	LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnIysPheCysGlnArg	301
DB	915	TTATTTTTTAAAGAAAGAAATCATCGAAGAAACAAAGGAGCAAAATAATCTGCCAGCGT	974
QY	302	TyrAspGlnLeuMetGluAlaLeuGluIysLysValGluArgIleGluAsnAsnProArg	321
DB	975	TATGATCAGCTCATGGAGGCGATGGGAGAAAAAAGTGACAGAAATAGAAAAATCTCTCG	1034
QY	322	ArgArgAlaIysGluSerIysValArgGluTyrTyrGluLysGlnPheProGluIleArg	341
DB	1035	AGGAAAGCTTAAAGAAAGCAAAACAAAGGGAATATCTATGAAAGCAGTTTCCAGAAATTCGA	1094
QY	342	LysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeu	361
DB	1095	AAACAAGAGAACAGCAAGAAAGATTTCAG---CGAGTTGGGCGAGGAGGAGCTGTCTT	1151
QY	362	SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu	381
DB	1152	TCAGCCACCATTTCTAGGAGTGCAGCATGAGATTTCTGAAATTTATTGATGGGCTCTCTGAG	1211
QY	382	GlnGluAsnLeuLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp	401
DB	1212	CAGGAGAAATTAATGAGAAACAAATTCGGCAGCTCTCTGTGATTCACCTATGATTTGAT	1271
QY	402	AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys	421
DB	1272	GCAGAACAAAGACGAGTCAAGTTCTAATACATGAAATGGGCTTATGGAGACCTATGAAA	1331
QY	422	ValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArg	441
DB	1332	GTGTATAAAGATAGGCAGTTTATGAATTTTGGACTGCACCATGAAAGGAGATCTTTAAG	1391
QY	442	GluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLys	461
DB	1392	GACAAAGTTTATCCAGCATCCAAAAAATCTTGGCAATTAATTCATCTTGGAGAGGAG	1451
QY	462	ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer	481
DB	1452	AGTGTCTCGATTGT	1511
QY	482	LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln	500

[illegible]

Db	5325	-----		-----	-CTGCCACCTGGCAGT	5342	
Qy	1854	HisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla	---	LeuGln	1872		
Db	5343	CATGATATGTTCCCTCCCTCCCT	---	TCAGTAAGAACTCAGGAGACCATGTTGCCAA	5399		
Qy	1873	GlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPro			1892		
Db	5400	CAGAGACCCAGTGTTCACAGGAACCAATGGAACCATGTAAATCACACCTTTGGATCCA			5459		
Qy	1993	SerLysProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThr			1912		
Db	5460	ACT	---	GCTCAGCTACGA	---	ATCATGCCA	5483
Qy	1913	PheProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeu			1932		
Db	5484	-----	CTGCCCTGGGGGC	-----		5498	
Qy	1933	MetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArg			1952		
Db	5498	-----		-----		5498	
Qy	1953	AlaAspThrGlyHisAlaPheLeuAlaIysProProAlaArgSerGlyLeuGluProAla			1972		
Db	5499	-----	CTTCAATAAGCCAAAGCCTG	-----		5519	
Qy	1973	SerSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAla			1992		
Db	5520	-----	CCAGCCTCCCGTTTACAC	5537			
Qy	1993	ThrIleAlaArgThrProAlaLysAsnLeuAlaProHisAlaSerProAspProPro			2012		
Db	5538	ACTGCTGGATGCTCCCTGGCTCTTGTG	-----		5567		
Qy	2013	AlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSer			2032		
Db	5568	-----	GATGCTGCAGCTTCTGCACCCAGATGATGTCTCCAAAACAAAGAGATATCT	5621			
Qy	2033	IleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerTyrSerProGluGly			2052		
Db	5622	TCTCACAGGTATGAACACCTAGC	-----	GATGCT	5651		
Qy	2053	ValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys			2072		
Db	5652	ATTGAGTGATACTCTGCGAGCTACCTCGGCCACCCAGGAGAACTGCAGACCTAT			5711		
Qy	2073	HisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly			2092		
Db	5712	CAGCCAGAGGTGTTAAGGCAATCAAGCGCAAAATGATCCTACGACAAATATGAAGGA			5771		
Qy	2093	ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer			2112		
Db	5772	CCA	-----	TTACATCACTATCGACCA	5795		
Qy	2113	GlnProSerSerProLeuLeuGlnThrAlaProGly	-----		2125		
Db	5796	CAGAAATCACCATCTCCCAACACACAGCTGCCCTCTTCACAGGCAGAGGAATGGGG			5855		
Qy	2126	ValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThr			2144		
Db	5856	CAAGTCCCGCAGGACCCATCGCTGATCATCACTTGCTGATCATCTCTGCAAAATATCACA			5915		
Qy	2145	GlnAspTyrThrArgHisHis	-----	ProGlnGlnLeuSerAlaProLeu	2159		
Db	5916	CAAGATTTTGTAGAATAACAGTTTCTCGCAGACTCCCCAGCAG	-----	CCTCCT	5966		
Qy	2160	ProAlaProLeuTyrSerPheProGlyAla	-----	SerCysProValLeuAspLeuArg	2177		
Db	5967	ACTTCTACATTCAGAACTACCTTCTGCTTGGTATCTACACTGTG	-----	AGG	6017		
Qy	2178	ArgProProSerAspLeuTyrLeuProProProAsp	-----	HisGlyAla	2192		

6018	ACTAAACATCAAAACCGTTACAGCCCAAGAAATCCAGGCTCAGTCTGTTCATCATCAAAAGA	6077
2193	Pro---AlaArgGlySerProHis-----SerGluGlyGlyLysArgSer	2206
6078	CCAGGTTCAAGGCTCTCTCCAGAAAACTTGTGGACAAATCCAGGGGAAGTAGGCGCTTGA	6137
2207	ProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSer	2226
6138	AAATCCCCAGAGAGAGAGTCAAGTCT-----TCTTCGGAGCGCTACGAGCCCATCTCC	6188
2227	ProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyr	2246
6189	CCACCCAG-----GTTCCGGTTGTGCATGAGAACACAGGACAGCTGTGCTGTCTTGTCT	6242
2247	ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer	2266
6243	CAGAGGGCGCAGAGCCTGCAGACGACAGAAATCATGCCCGCTCACCAGGGAGTATAGC	6302
2267	GlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLys	2286
6303	TACTTGCCTTCATTCTTCCAAAGCTT---GAAATACATCACCCATGGTTAAATCAAAG	6359
2287	LysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIle	2306
6360	AAGCAGAGAGATTTTCGTAAGTTGAATCTCTGTGGAGTGACTCTGATGGCAGCT	6419
2307	SerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThr	2326
6420	GCTCAGCAGCAGAACTGAGATCTTTAATCTGCCAGCAGTTACTACGTACAGCTCAGTTAGC	6479
2327	TyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIle	2346
6480	TCTAGAGGCCATTCTTTGTGTGATCTCTCCAGT---AATCTGGCGTGAAGACATTATC	6536
2347	ArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu-----Ser	2360
6537	AGGAAGCTCTCATGGGAAGCTTTGATGACAAAGTTGAGGATCATGAGTTGTCATGTCC	6596
2361	ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla	2380
6597	CAGCTATGGGAGTA-----GTCCCTGTGTTACTGCCAACACCTCAGTT-----	6638
2381	MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly	2400
6639	-----GTGACCAGTGTGTGAGACACGAAGAGAGAGAGGGGACCCATCACCTCATTCAGGA	6692
2401	Gly-----LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAla	2418
6693	GGAGTTTGCAAAACCAAGCTGTATCAGCAAGCTAAACAGCAGGAAATCTAAGTCTCTTATA	6752
2419	Pro-----GlyLeuAlaSerGlyAspArgProProSerValSerValHisSerGlu	2436
6753	CCTGGCAAGGCTACTTAGGAACGGAACGGCCCTTTCAGTCTCTCTGTACATTTCAGAA	6812
2437	GlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSer	2456
6813	GGGGATTACCATAGGCAGACGCCA-----GGGTGGGCGCTGGGAAGACAGCGCCCTCTTCA	6866
2457	AlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMet	2476
6867	ACAGGCTCAACTCAGTTTTCCTTTAATACCTCTGACTATGGCGATG-----CTCAGC	6917
2477	AlaSerProProProGlyLeuProAlaGlySerGly---ProLeuAlaGlyProHis	2495
6918	AGTACTCCACCAACACCGATTGATGTGCTCCCTCTCGCGGTGAACCAAGCAGCTCTCTCAC	6977
2496	-----HisAlaTrpAspGluGluProLysPheProLeuLeuCysSerGlnTyrGluThr	2512
6978	CAACAGAACAGGATCTGGGACGAGAGCCTGCCCACTGCTCTCAGCACAGTACGAGACC	7037
2513	LeuSerAspSerGlu	2517
7038	CTGTCCGATAGTATGAT	7052

QY 265 nTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLy 285
DB 938 GTACCAACGAGAAATCATCAAGCAACACCGAGTGTGATGAGGAAAAAATCTATTTTATTTTAA 997
QY 285 sArgArgAsnHisAlaArgLysGlnTTrpLysGlnLysPheCysGlnArgTyrAspGlnLe 305
DB 998 AAGAGAAATCATGCAAGAAACAAAGGGAACAAATAATCTGCCAAGCTTATGATCAGCT 1057
QY 305 uMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgAlaLy 325
DB 1058 CATGGAAGCATGGAGAAAAAGTGGACAGAAATAGAAAAATATCTCGGAGGAAAGCTAA 1117
QY 325 sGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgG1 345
DB 1118 AGAAAGCAAAACGAGGAATATCTATGAAAAACGATTTTCCAGAAATTCGAAAAACAAAGAGA 1177
QY 345 uLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAl 365
DB 1178 ACAGCAAGAAAGATTTCAG---CGAGTTGTCTAGAGGGAGCTGTCTTTACGCCACCAT 1234
QY 365 aAlaArgSerGluHisGluValSerGluIleLeuAspGlyLeuSerGluGlnGluAsnLe 385
DB 1235 TGCTAGGAGTGAGCATGAGATTCTCGAAATTAATTGATGGTCTTCTGAAACAGGAGATAA 1294
QY 385 uGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnG1 405
DB 1295 TGAGAAAGCAAAATCGTCAGCTTTCTGTGATTCACCTATGATGTTTGTGATGAGAAACAAAG 1354
QY 405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAs 425
DB 1355 AAGGGTCAAAATTCATCAATATGATGGCTGTGAGGAGATCCCAATGAAGGTTTATATAAGA 1414
QY 425 pArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPheArgGluLysPheMe 445
DB 1415 CAGACAGTTTATGATGTTTGGACTGACCATGAAAGGAGATCTTTAAGGACAGTTTAT 1474
QY 445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaG1 465
DB 1475 CCAGCATCCAAAAAATCTTGGACTTAATTGCATCTTATTTGGAAGGAAGAGTGTCTCTGA 1534
QY 465 uCysValLeuTyrTyrTrpLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgAr 485
DB 1535 TTGTGTTTATATTACTATTATTAACCAAGAAAAATGAGAATTATAAGGCCCTCTGTAGAAG 1594
QY 485 gSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnG1 504
DB 1595 GAATTATGGAACACGACAGCGCAGAAATCAGCAGATTGCCCGTCCCTCACAAGAGAAAA 1654
QY 504 nGlnGlnGlnGlnGlnProMetProArgSerSerGlnGlnGlnGlnLysAspGluLysG1 524
DB 1655 AGTAGAAGAAAAAG-----GAAGAGATAAAGCAGAAAAAAC 1690
QY 524 uLysGluLysGluAlaGluLys-----GluGluGluLysProGluValGluAsnAspLy 542
DB 1691 AGAGAAAAAGGAAGAAAGAAAGGATGATGATGAAGAAAAAGATGAAGGAAGCACTCTAA 1750
QY 542 sGluAspLeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysG1 562
DB 1751 AGAAACAAACCAAGAAAAAGACAGAACGGAAGCCACAGCAGAGAACCTGAAGAGAGA 1810
QY 562 uAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlnArgArgLysGlyArgI1 582
DB 1811 GCAGGTCACTCCAAAGGGGGGAAAGACTGTCTAACAGCCAAAGGCCCGGGAAGGGCGGGT 1870
QY 582 eThrArgSerMetAlaAsnGlu-----AlaAsnSerGluGluAlaIleThrProG1 599
DB 1871 CACCAGGTTCGATCAACAGTGAAGCTGCAGCTGCGCAATGCTGTGCTGCACGCCACTGAGGA 1930
QY 599 nGlnSerAlaGluLeu-----AlaSerMetGluLeuAsnGluSerSe 613
DB 1931 GCCCCCGCCACCCCTCGCCGCCACACACAGAGCCCATTTCTACAGAACCTGTTCGAGACTTC 1990
QY 613 rArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuHisGlyArgAs 633

DB 1991 AAGATGGACAGAAAGAAATGGAAGTTGCTAAAAAAGGCGCTGTGAGAAATGCTGTGATAA 2050
QY 633 nTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTy 653
DB 2051 CTGGGAGCCATTGCTAAATGTTGGAACTAAAGTGAAGCCAGCTGTCGCAAAACTTCTA 2110
QY 653 rPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMe 673
DB 2111 TTTTAACTATAAAAGACGGCATTAATCTTGACAACTTTTGCAGCAACATAAAGAGAAAGC 2170
QY 673 tGluLysGluArgAsnAlaArgArgLysLysLysLysAlaProAlaAlaAlaSerGluG1 693
DB 2171 TTCACGGAACCCCGTGAGGAACGAGATGATCTCAGTGTGAAAGTGTCTCTCCACTGT 2230
QY 693 uAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyValSerGlyAs 713
DB 2231 TTCCTGCC-----CAAGAGATGAAGATATTGAAGCTCA-----AA 2266
QY 713 nGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProAr 733
DB 2267 TGAGGAGGAAAAATCCAGAGATAGTGAAGGT----- 2297
QY 733 sGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSe 753
DB 2298 -----GCTGAAAAATAGTTCTGTATACAGAAAGTGTCTCCCTC 2332
QY 753 rProHis---ThrGluAlaAlaLys-----AspThrGlyGlnAsnGlyProLysPr 769
DB 2333 TCCTTCACAGTTGAAGCTGCCAAGTCCAGTGAAGACAGCAGCGTGAATAATGCT----- 2384
QY 769 oProAlaThrLeuGlyAlaAspGlyProProGlyProProThrProProArgArgTh 789
DB 2385 ----- 2386
QY 789 rSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProPr 809
DB 2387 TTCTCGAGAAACACCGAGCCTGTGCTGAGCTTGAGGCCACCACCTGACCCCTGCACCCTG 2446
QY 809 oProAlaProProSerProSerAlaProProValValProLysGluGluLysGluG1 829
DB 2447 TGATCTCCC-----TCCTTCAGCTTCCACCAACCAACAGAGAAAGGAAAGCGT 2500
QY 829 uGlu-----ThrAlaAlaAlaProProValGluGluGlyG1 841
DB 2501 GGAGGCCAGGTGACCCGACAGCGCCAGTGCAGAGACCGCAGACCGATGCGACGTAGACCA 2560
QY 841 uGluGln-----LysProProAlaAlaGluGluLeuAl 852
DB 2561 TGAGGAGTGCCTGCGCGAGGCGAGTCTGTCTTGTATCCACAGCCCTTACCACCA----- 2615
QY 852 aValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluG1 872
DB 2616 -GCCGACTCCGTGGACCCAGAAATGAGGTA---CCAGAAAAATCTGCGTCTAAAGGTGA 2671
QY 872 uGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLy 892
DB 2672 AGGGGATGCCAAGAAAGAGACTTGGAGAGACAC----- 2705
QY 892 sAlaGluLysLysGluGlyCysSerGlyArgAlaThrThrAlaLysSerSerGlyAlaPr 912
DB 2706 -AGTGAGAAGACAGAGCGCTAGAGATGAAGACGCTGGTAGTGGCTGAACACAGATCGAGAGGCC 2764
QY 912 oGln-----AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAl 929
DB 2765 TGAGCCACAGTTCAGACGACACTCCAGTGCCTTGCAGTGCAGATGAGGGTGTGGATGG 2824
QY 929 aGluGlyGlyAspLysAsnArgLeu-----LeuSerProArgProSerLeuLeuThrPr 947
DB 2825 AGAG---CCAGAGAGCAGAGAGTGTTCCTCATGGATGCAAGACCTTCATTTGTTACTCC 2881
QY 947 oThrGlyAspProArgAlaAsnAlaSerProGlnLysPro-----LeuAspLeuLysG1 965

Qy	1621	laePLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyI	1641
Db	4808	AGAGGTTTATCGGAGCCACCTTACCACGACCTTGGATCCA---GCTATGCGCTTTACAG	4864
Qy	1641	eProLeuAspAlaalaalaIaTyrLeuProArgHisLeuAlaProAsnProThrTy	1661
Db	4865	GGCTTTGGATCCTGCTGCTTACCTTGTGTACAGAGACAGCTTTTCAACAACCCGAGNTA	4924
Qy	1661	rProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaalaLeuGluAs	1681
Db	4925	CCCAAGTCAGTACCAGCTCTAT-----GCAATGGAGAA	4957
Qy	1681	n---ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisenThrAl	1700
Db	4958	TACAAGGCAGACAATCTCTCAACGATTACATTTACCTCACAGCAGATGCAGTGAATCTG--	5015
Qy	1700	aThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerIe	1720
Db	5016	-----CGCCCTGATGTCAACAGGGGACTGTGCCACAGAGCAGAGCCGCACT	5059
Qy	1720	uAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLe	1740
Db	5060	GGGCTCCCTTACCCAGCT---ACAAGAGGATCATTTGACTGACCAATATGCTCCA--	5114
Qy	1740	uProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTy	1760
Db	5115	-ACAATCTTAGTGCTCATGCAGGGGAAACGAGCACCCCTCCCATGACAGGATCAGTA	5173
Qy	1760	rLeuProThrAlaProGlnProPheSerSerArg---HisSerSerSerProLeuSerPr	1779
Db	5174	TATTCCCTGGTACACAGGTTACTTCCCTCCAGGCCATATAAGCGTCTTCTCTGTCTCC	5233
Qy	1779	oGlyGlyProThrHisLeuThrLysProThrThrThrSerSerGluArgGluArgAs	1799
Db	5234	AGGACCCCAACACACTT-----GCAGCAGCTGCAGTGTCTGAGAGGAAACGAGA	5284
Qy	1799	pArgAspArgGluArgAspArgAspArgGluArgGlySerIleLeuThrSerThrTh	1819
Db	5285	ACGGGAAGGGAGAGAGGCGGAAACGTGAGCGCGAGCTGAGCGTGAACGTGAACGCGA	5344
Qy	1819	rThrValGluHisAlaPro-----IleTrpArgProGlyThrGluGlnSerSerGl	1836
Db	5345	AAGGATCGCTGCTGCTCCGCTGACCTCTACCTACGACCAAGGTTCAAGAACAG--	5396
Qy	1836	ySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHi	1856
Db	5397	-----CCAGGGCGCCCTGGCAGCCACGGNTA	5422
Qy	1856	sAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArgPr	1875
Db	5423	TGTTTCGCTCCCTTCCGCT---TCAGTAAAGAACTCAGGAGACCATCTCTGCAACAGAGACC	5479
Qy	1875	oSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysPr	1895
Db	5480	CAGTGTTTTCAGGGCCCAATGAAACGAGTGAATCACACCTTTGGACCCCACTGCTCA	5539
Qy	1895	oThrValLeuArgSerThrSerThrSerSerProVal-----ArgProAlaAl	1911
Db	5540	GCTACGATCATGCGACTGCTTCTGGGGCCCTTCCATAAGTCAAGCCCTGCAGCCTC	5599
Qy	1911	aThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProTh	1931
Db	5600	CCGTTACAACACTGCTCGGATGCC---CTGGCTGTCTTGTGTGCTGCTGCAGCTTCTGC	5656
Qy	1931	rLeuMetGluProValLeuLeuProLysGluAlaProArg-----	1944
Db	5657	ACCCAGATGGATGTTTCCAAAACAAAGAGAGTAAGCATGAAGCTGCCAGGTTAGAAGA	5716
Qy	1945	-----ValAlaArgProGluuArgProAlaAlaSerSerProSerLysGl	1958
Db	5717	AAATTTCAGAGCAGGTCAGCAGCATGTAGTGAACACAGCAGCTAGACAGAAAAAACCCT	5776
Qy	1958	aPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGl	1978

Db	5777	GGAGGTGGGAAGAAGATCTGTTTCAGTGTGTGTGCACCTTCTTCAGCCCTCTCCAAATGGGCAA	5833
Qy	1978	ySerGluProArgProLeuVal	1987
Db	5837	GGCCACGCTCATGCGCTCAGTAGTGATTTCTGAGGCTGGGAAAGATAAAGGGGCTCTCTCC	5896
Qy	1987	oValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn	2005
Db	5897	AAATCCAGATATCAGGAAGAGCTAAGGACCCGAGGGAAGACTACCATTTACTGCAGCTAA	5956
Qy	2005	sHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLys	2025
Db	5957	CTTCATAGACGTGATCATACCCCGGCAAAATGCTCGACAAGAGTCGAGGGAACCTGG	6016
Qy	2025	sThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGln	2045
Db	6017	CTCTCAAAGTTCAGACTTTCT	6058
Qy	2045	ySerSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThr	2065
Db	6059	GTATGAACCGGCTAGTATGCCATTCAGGTGATAAGTCCCGCAGCTCCTGCACCAACC	6118
Qy	2065	rHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGln	2085
Db	6119	CCAGGAAGGCCACGAGCCCTATCAGCCAGCATGTGTTAAGCAATCAACAGCAAAATGA	6178
Qy	2085	uLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProHis	2105
Db	6179	GTCCACTCGACAGTATGAAGGTCCA	6211
Qy	2105	sLeuArg	2120
Db	6212	TTATCGGTCCACGAGGAATCACCATCTCCACAGCAACAGCCACCATGCCCCCATCTTC	6271
Qy	2120	uGlnThr	2138
Db	6272	CCAGTCAGAGGGAATGGACAGGTGCCAGGACCCATCGACTCATCATCTTGTGTGACCA	6331
Qy	2138	sIleSerGluValIleThrGlnAspTyrThrArgHisHis	2157
Db	6332	CATCTGTCAAAATTATCACACAAGATTTTGTAGAAAATCAAGTTCCCTCGCAGGGTTCTAC	6391
Qy	2157	aProLeu	2173
Db	6392	TTCTACATTCCAAATTCACCATCTGTTTGTTCATCCACACT	6434
Qy	2173	lLeuAspLeuArgProProSerAspLeuTyrLeuProProAsp	2189
Db	6435	-----GTAAGAATAAAACCTCAAGCCCTACAGCCAGATCATCAGTCTCAGACTGT	6487
Qy	2190	-----HisGlyAlaProAla	2202
Db	6488	CTTGATCCGACACCAAGTCTTAGAGTCTCTCCAGAAAATCTTGTGATATAATCCCGGG	6547
Qy	2202	yGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIle	2222
Db	6548	AAGCAGGCTGGAAAATCTCCAGAGGAGTCTATATC	6598
Qy	2222	eGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTy	2242
Db	6599	TGAGCCCATCTCCCAACCCCAAGGCCCTGCTGTG	6652
Qy	2242	rProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetClySerLysSerPr	2262
Db	6653	GTTGCTCTTGTACAGAGGGAGTGGACCTGCTGAGCAAAAGAGTGATTCTCGATCACC	6712
Qy	2262	oGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMe	2282
Db	6713	AGGAGTATAAGCTACTTGCCTTCATCTTCACCAAGCTT	6769
Qy	2282	tValLysSerLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluPr	2302

QY 302 TyrAspGlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsnProArg 321
DB 1171 TATGATCAGTCTATGGAGGATGGGAGAAAAAGTGACAGAAATAGAAAAATATCTCTCG 1230
QY 322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArg 341
DB 1231 AGGAAGCTTAAGNAGCAAAACNAGGGAATACTATGAAAGCAGTTTCCAGAAATTCGA 1290
QY 342 LysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeu 361
DB 1291 AAACAAAGAGAACAGCAAGAAAGATTTTCAG--CGAGTTGGCGAGAGGGAGCTGTCCT 1347
QY 362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleLeuLeuAspGlyLeuSerGlu 381
DB 1348 TCAGCCACCATTGCTAGGAGTGCAGCATGAGATTCTGAAATATTATGATGGCTCTCTGAG 1407
QY 382 GlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp 401
DB 1408 CAGGAGAAATATGAGAAACAAATGCGGCAGCTCTCTGTGATTCACCCTATGATGTTGAT 1467
QY 402 AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421
DB 1468 GCAGAACAAAGACGAGTCAGTTTCATTAACATGATGGCTTATGGAGGCCCTATGAAA 1527
QY 422 ValTyrLysAspArgGlnValMetAsnMetTrpSerGlnGlnGluLysGluThrPheArg 441
DB 1528 GTGTATAAGATAGCGCAGTTTATCAATGTTTGGAGTACCCTGACCATGAAAGAGATCTTAAAG 1587
QY 442 GluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLys 461
DB 1588 GACAAGTTTATCCAGCATCCAAAAAATTTGGACTAATTCATCATCTACTTTGGAGAGGAAG 1647
QY 462 ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer 481
DB 1648 AGTGTTCTGATGTGTTTGTATTACTATTATTAACCAAGAAAAATGAGATTTAAAGCC 1707
QY 482 LeuValArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
DB 1708 CTGCTCAGAGGAATATGAGAAACGAGAGGAGAAACAGCAAAATTTGTCGACCTCG 1767
QY 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
DB 1768 CAAGAGAAAAAGTAGAGAAAAAGAGAG--GATAAAGCAGAAAAAAGAGAAAAA 1821
QY 521 AspGluLysGluLysGluLysGluAlaGluLysGluLysProGluValGluAsn 540
DB 1822 AAGAGAGAAAAAGAGATGAGAGAAAAAGATGAAAGAAAGAGACTCAAGAAAAAT 1881
QY 541 AspLysGluAspLeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGlu 560
DB 1882 ACCAAGGAA-----AAGGACAAGATAGATTGTTACAGCA--GAAGAAACTGAGGAA 1929
QY 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGly 580
DB 1930 AGAGAGCAAGCCACACCCCGGGCGGAGAAAGACTGCCAACAGTCCAGGCGCGCTAAGGGC 1989
QY 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluAlaIleThrProGlnGln 600
DB 1990 CGGATCACAGGTCATGACAAACGAAAGCTGCAGCTGCCAGTCTGCAGCCGCGAGCGCT 2049
QY 601 Ser-----AlaGluLeuAlaSerMetGluLeuAsn 610
DB 2050 ACTGAAGAGCCCCCACCACCTCTGCCACCGCCACAGACCCATTTCTACAGAGCCTGTG 2109
QY 611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuHis 630
DB 2110 GAGACCTCTCGATGGACAGAGAGAAAAAGGAGTTGCTAAAAAAGGCTTAGTAGAACAT 2169
QY 631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
DB 2170 GGTCTGAATCTGGGCAATTTCTAAATTTGTTGGGAGCAAAAGTGAAGCTCAATGTAA 2229
QY 651 AsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnHisLys 670

DB 2230 AACTTCTATTATTAACATATAAAGCGCAGACAATCTTGACAACCTCTTACAGCAGATAAA 2289
QY 671 LeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAlaAla 690
DB 2290 CAGAAAACTTCAGAAAAACCTCGTGAAGACGAGATGTGTCTCAATGTGAAGTGTGCT 2349
QY 691 SerGluGluAlaAlaPheProProValValGluAspGluMetGluAlaSerGlyVal 710
DB 2350 TCCACTGTTTCTGCT-----CAGGAGGATGAAGATATTGAAGCTCC----- 2391
QY 711 SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
DB 2392 -----AATGAAGAAGAAAAATCCAGAAGACAGAGGAA----- 2421
QY 731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
DB 2421 ----- 2421
QY 751 IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPro 770
DB 2422 -----GTTGAAGCTGTCAAG----- 2436
QY 771 AlaThrLeuGlyAlaAspGlyProProGlyProProGlyProProThrProProArgThrSer 790
DB 2437 -----CCAGCGAGACAGCTCTGAAAAATGTACTTCT 2469
QY 791 ArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro 810
DB 2470 CGAGGAAACACAGAACTCGCGTTGAGTTGAG-----CCACACCGGAAACT 2517
QY 811 AlaPro---ProSerProSer---AlaProProProValValProLysGluLysGlu 828
DB 2518 GCACCCAGTACATCTCCTCTCTAGCAGTTCCAAGTACAAAAACCGCTGAAGTGAAGT 2577
QY 829 GluGluThrAlaAlaAlaProVal----- 837
DB 2578 GTGAGACCCAGGTGAATGACAGCATCAGTCTGAGACAGCAGACAGATGGATGTAGAT 2637
QY 838 -----GluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu 851
DB 2638 CAGCAGAGACAGTCTGAAGAGGGTCTGTTTGTGATCCCCCAGCCGCTACCAAA--- 2694
QY 852 AlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGlu 871
DB 2695 ---GCTGACTCTGAGACGTTGAAGTGAAGTCCAGGGTCCAGAAAAACCATGCATCTAAAGTTGAA 2751
QY 872 GluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeu 891
DB 2752 GGTGATAATACCAAGAAAGAGACTTGGATAGAGCC----- 2787
QY 892 LysAlaGluLysLysGluGlySerGlyArgAlaThrThrAlaLysSerSerGlyAla 911
DB 2788 ---AGTGAAGAGGTGAACCTGAGATGAAGATTGGTGGTAGCTCAGCAAAATAATATGCC 2844
QY 912 -----ProGln---AspSerAspSerSerAlaThrCysSerAlaAspGluVal 926
DB 2845 CAAAGCCCGAGCCCGCCAGTCAGACATGATTCCAGTCCACGTCGAGCGCTGATGAG--- 2901
QY 927 AspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro-----ArgProSer 943
DB 2902 ---GATGTGATGAGAGCGCAGAGAGGAGAGAAATGTTTCTATGAGACTCAAGACCTTCA 2958
QY 944 LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeu 961
DB 2959 CTGTAAACCCCACTGGATCTATCTACGTCG---TCATCTCCGTTTAAACCAAAATCCACTG 3015
QY 962 AspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProIle----- 976
DB 3016 GATCTGCCACAGCTTCAGCATCGAGCTGTGTTATCCCAATGTTATCTCTGCACCCCA 3075
QY 977 -----GlnValThr 979
:::

Db 3076 TGTAAATACCAATTTGAAACCCAGTGGCGGTATGCTCTCTACAGGACACATTTAA 3135
Qy 980 LysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPro 999
Db 3136 GCAATGCATGAGTCAGCACTCTCTGGAG----- 3162
Qy 1000 ProProGlnAsnLeuGlnProGluSerAspAlaProGlnProGlySerPro 1019
Db 3163 ---GAGCAGCGGAGAGACAGACAGATAGATTTGGAATGTAGAGTTCTCAAGTCCA 3219
Qy 1020 ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaGluAla 1039
Db 3220 TGTGGGCATCTCAAGAGTCCA-----AACAGAGAG----- 3249
Qy 1040 GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPro 1059
Db 3250 -----TGG----- 3252
Qy 1060 ArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaPro 1079
Db 3253 ---GAAGTCTCTCAGCGCTGCTCACAT-----CAATTGATACTATCTCCCT 3297
Qy 1080 ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro 1099
Db 3298 GAAGCGTTCCGGCTTCOG-----ACAACTCGACCAACAGGCCACCGCC 3342
Qy 1100 ProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGlu 1119
Db 3343 CCTCTCATCCCGTCATCCAAACACACAGTGGCTTCAGAAAA---CCATCTTTTATA--- 3396
Qy 1120 ArgGluIleGlyAlaIleSerGlnGly-----MetSerValGlnLeuHis 1134
Db 3397 ---ATGGAGGCTCCATCTCACAGGGAACACAGGCACCTATTGTACTCTCATATCAG 3453
Qy 1135 ValProTyrSerGluHisAlaLysAlaPro---ValGlyProValThrMetGlyLeuPro 1153
Db 3454 GCTTCTTACACTCAAGAAACACCAAGCGCTCAGTAGGACTATCTCTCTGGACTGCCA 3513
Qy 1154 LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluLeuSer 1173
Db 3514 CGCAACAGGAATCTGCCAAATCAGCTACTTTCCTTACATCAAGCAGGAAGAAATTTCT 3573
Qy 1174 ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSer 1193
Db 3574 CCGCAAGCCAAACTCACAACCTGAGGCTCTGTGGTC---AGGGCCCAACATGAAGT 3630
Qy 1194 ValLeuArgGlyThrAlaLeuGlySerValProGlySerIleThrLysGlyIlePro 1213
Db 3631 GTAGTCAGAGTACCGCA---GGAGCCATCAAGAGGAAGTATACTCGGGAACTCCA 3687
Qy 1214 SerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThr 1233
Db 3688 ACCAGCAAAATTTTCAGTGGAGAGCATCTCCATCCCTACCGGGCTCTATCACTCAGGGCACC 3747
Qy 1234 ProAla-----AspValLeuTyrLysGlyThrIleThrArg 1245
Db 3748 CCGGCTCTGCCCGACCTGCATACCAACAGAGCCTTGTGAGGGTCCATTCGAGA 3807
Qy 1246 IleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLys 1265
Db 3808 ATGCCCATTTGAAGACAGCAGTCTCT-----GAGAAAGCCAGAGGAAGTGCATCCAA 3861
Qy 1266 GlyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
Db 3862 GGCCATGTTATTTATGAAGCAAAAGTGACATATCTTGTCATATGATATATATTAAGAT 3921
Qy 1286 ValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAla 1305
Db 3922 GCC-----CGAAGAGGACTAGGAGTCCAGACAGCTCATGAATCAGT 3966
Qy 1306 AlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer----- 1323
Db 3967 TTA---AAGAGAGCTATGAATCAGTGGAAAGGAATATAAAGCAAGGATGTCAATGAGG 4023

Qy 1324 -----SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg 1339
Db 4024 GAGTCTCTGTATCAGCACCTGTAGAGGGCTGATATCCCGAGCATTACCC-----AGG 4077
Qy 1340 HisSerProHis---HisLeuLysGluGlnHisIleArgGlySerIleThrGlnGly 1358
Db 4078 GGGAGTCTCTCATTTCTGACCTCAAAAGAAAGAGTGTATTGTCTGGCTCCATAATGACGGG 4137
Qy 1359 IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaLysLeuLeu 1378
Db 4138 ACACCAAGAGCAACAACTGAAAGCTTTGAAGATGGCTT---AAATATCCCAACAAAT 4194
Qy 1379 LysArgGluGlyThrProProProProProSerArgAspLeuThrGluAlaTyrLys 1398
Db 4195 AAAAGGGAAGT-----CCTCCATACAGCATTTGAAGTGGCCATT--- 4236
Qy 1399 ThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrVal 1418
Db 4237 -----ACCAAGGAAACCA---TATGATGGC---ATCACCACCATC 4272
Qy 1419 LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeu----- 1433
Db 4273 AAAGAAATGGGGCGTTCATTCATGAGATTCAGAGCAAGATATTTTAACTCAGAAAGT 4332
Qy 1434 ArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGln 1453
Db 4333 CGAAATCTCCAGAGTGGTCCAGAGCACACGGCCGATAATTGAGGGTTCATTTCACAG 4392
Qy 1454 GlyThrProLeuLysTyrAspThrGlyAlaSerThrGlySerLysLysHisAspVal 1473
Db 4393 GGCACACCAATAAAGTTTGACAAAC---TCAGGTCAATCTGCCATCAACACACATGTC 4449
Qy 1474 ArgSerLeuIleGlySerProGlyArgThrPheProValHisProLeuAspValMet 1493
Db 4450 AAATCTTAAATCAGGGGCTTAGCAAACTATCCCTGGAAATCCCTCGCTGGAAATGTG 4509
Qy 1494 AlaAsp---AlaArgAlaLeuGluArgAlaCysTyrGlu-----GluSer 1507
Db 4510 CCAGAGACATAAAGTGTAGAACCGGAAATATGAGATGTGAAAGCAGCGGAGACC 4569
Qy 1508 LeuLysSerArgProGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAlaPro 1527
Db 4570 GTGCGTTCGGGCACACACAGTGTGTAAGCTCTGCCCTCGCTTCTTATGATGACACAGT 4626
Qy 1528 ValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGly 1547
Db 4627 -----CTGCATGAAGCTCCCAAGACACACTGAGCCCTGGGATTTATGATGACACAGT 4680
Qy 1548 Ala-----ProPheAlaGlyHisLeuProArgGlySerProValThrMet 1562
Db 4681 GCACGGAGGACCCCTGTGAGTTATCAAAACACCATGTCCAGAGGCTCACCCATGATGAAC 4740
Qy 1563 ArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAsp 1582
Db 4741 AGAACTTCTGATGTTACAAATTCCT-----CCTAAACAAGTCTACCAATCATGAA 4788
Qy 1583 ArgLysLeuThrSerThrProArgGlu-----IleAlaLysSerProHisSer 1598
Db 4789 AGAAATCCGACTGACCCCTACCCAGGGAAGTATCCAGCGAAGTCTCCAGTGCCT 4848
Qy 1599 ThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyVal 1618
Db 4849 GGGGTGGACCCCTGCTCGTGAGCCAC-----AGTCCGTTTGTATCCCATCACAGAGCAGC 4902
Qy 1619 SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrIlePro 1638
Db 4903 ACTGAGGCGGAGGTTTATTTGGAGCCACCTGCCCCACCAATTTGGATCCA---GCCATGCCT 4959
Qy 1639 ArgGlyIleProLeuAsp---AlaAlaAlaIleTyrLeuProArgHisLeuAlaPro 1657
Db 4960 TTTCAGAGGCTTTGGATCTCGAGGGCTGCTTACTCTGTTTCAGAGACAGCTTTCACCA 5019

Db 469 CCAAGTTCTCAGAGCCCTCAAGAA-----AGGAGAACTAGTTATGAAACCGTTTCATCCA 522
Qy 94 GluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeu 113
Db 523 GGCCCATCCCACTGGATCATGATTCATCGAATCGAAGCGACCATGCTGGACAGGTT 582
Qy 114 ProAsp-----ProLeuLeuArgProSerPro 122
Db 583 TCTGATCTCATTTTCAGCGTGTGCTGCTCGCGTTTTCGCTTTAGTGCAACCGCTGCCA 642
Qy 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db 643 -----GAAGGGCTGAGGGCT---TCTGCAGATGCTAAGAGGATCCAGCATTCGGA 690
Qy 143 GlyLysLeuGlu---ProValSerProProSerProProHisThrAspProGluLeuGlu 161
Db 691 GGCACATGAGCTCCATCTCTCCATTTTCGGGGCAACCATGTGGAGATGATCAAAAT 750
Qy 162 LeuValProProArgLeuSerLysGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 181
Db 751 GCTTCACCTTCAAACTCTCAAGGAGAGCTTAATACAGAGTATGGATCGTGTAGATCGA 810
Qy 182 GluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeuGlu 201
Db 811 GAAATTCGCAAAAGTAGAACAGCATCTTAAATCTGAAAGAGAAACAAACACAGCTTGA 870
Qy 202 GluGluAlaLysProGluProGluLysProValSerProProIleGluSer 221
Db 871 GAAGAGCAGCTAAACCTCTGAGCCCTGAGAACCCGCTGCTCCCTCTCTGTGGAGCAG 930
Qy 222 LysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaA 241
Db 931 AAACCGCAGTATTGTCCAATTAATTATGATGAGATCGGAAAGAGAGAGCT 990
Qy 242 HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
Db 991 CATAAATTTTGAAGGTCTTGGCCCAAAAGTTGAATGCTGCACTGTATTAACGACCATCA 1050
Qy 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
Db 1051 GATACCAAGGTGTACCATGAGACATCAAGACAAACAGGTGATGAGGAAAGAACTCAIT 1110
Qy 282 LeuTyrPheLysArgArgAsnHisAlaAtrLysGlnTyrLysGlnLysPheCysGlnArg 301
Db 1111 TTATTTTAAAGAGAAATCATGCAAGAAACAAAGGGAACAAAGAAATCTGCCAGCT 1170
Qy 302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArg 321
Db 1171 TATGATCAGCTCATGGAGGATGGGAGAAAGAGTGGACAGAAATAGAAATAATCTCTCG 1230
Qy 322 ArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArg 341
Db 1231 AGAAAGCTAAAGAAAGCAAAACAGGGAATCTATGAAAGAGAGTTTCCAGAAATTCGA 1290
Qy 342 LysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeu 361
Db 1291 AAACAAGAGAAACAGCAAGAAAGATTTCAG--CGAGTTGGGAGAGGGAGCTGCTTT 1347
Qy 362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu 381
Db 1348 TCAGCCACCATTCAGGAGTGAAGATGAGATTTCTGAATTTATGATGGCTCTCTGAG 1407
Qy 382 GlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAsp 401
Db 1408 CAGAGAAATTAATGAGAAACAAATGCGGCAGCTCTCTGTGATTCACCTATGATTTTGTAT 1467
Qy 402 AlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys 421
Db 1468 GCAGAACAAAGACGAGTCAGTTTCATTAACATGATGGGCTTATGGAGGCCCTATGAAA 1527
Qy 422 ValTyrAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThrPheArg 441

Db 1528 GTGTATAAGATAGGACAGTTTATGAATGTTTGGACTGACCATGAAAGAGAGATCTTTAAG 1587
Qy 442 GluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLys 461
Db 1588 GACAAGTTTATCCAGCATCCAAAACCTTTGGACTAATTCATCATCATCTTTGGAGAGAG 1647
Qy 462 ThrValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLysSer 481
Db 1648 AGTGTTCCTGATGTGTTTGTATTACTATTAAACCAAGAAAAATGAGAATTATAAGCC 1707
Qy 482 LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
Db 1708 CTGTCAGAGGAATTTATGGAAACCGCAGAGGAGAAACCCAGCAAAATTCGACCCCTCG 1767
Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLys 520
Db 1768 CAAGAAGAAAAAGTAGAAGAAAAAGAGAG-----GATAAAGCAGAAAAACAGAAAAA 1821
Qy 521 AspGluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsn 540
Db 1822 AAAGAAGAAAAAGAGATGAAGAGAAAAAGATGAAAAAGAGACTCCAAAGAAAAAT 1881
Qy 541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGlu 560
Db 1882 ACCAAGGAA-----AAGGACAAAGATAGATGGTACAGCA--GAAGAACTGAGGAA 1929
Qy 561 LysGluAlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgLysGly 580
Db 1930 AGAGACCAAGCCACACCCCGGGGCGAAGAGCTGCCAACAGTCAGGGCCCGCTAAGGGC 1989
Qy 581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln 600
Db 1990 CGGATCACCAGGTCCATGACAAACAGAGCTGCAGCTGCCAGTCTGCAGCGCAGCGCT 2049
Qy 601 Ser-----AlaGluLeuAlaSerMetGluLeuAsn 610
Db 2050 ACTGAAGAGCCCCACCATCTCTGCCACCGCCACCAACCCATTTCTACAGAGCTGTG 2109
Qy 611 GluSerSerArgTyrThrGluGluGluMetGluThrAlaLysLysLysLeuLeuGluHis 630
Db 2110 GAGACCTCTCGATGGACAGAGAGAAAGAAATGGAAGTTGCTTAAAAAGGTCTAGTAGAACAT 2169
Qy 631 GlyArgAsnTyrSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
Db 2170 GGTGCTAACTGGCAGCAATTTCTAAATGGTGGGAACGAAAGTGAAGCTCAATGTAAA 2229
Qy 651 AsnPheTyrPheAsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLys 670
Db 2230 AACTTCTATTTTAACTATAAAGCGGCACACAAATCTTGACAACTCTTACAGCAGCATAAA 2289
Qy 671 LeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAlaA 690
Db 2290 CAGAAAACCTTCAGAAAACCTCGTGAAGAGCGAGATGTGTCTCAATGTGAAGTGTGCT 2349
Qy 691 SerGluGluAlaAlaPheProValValGluAspGluGluMetGluAlaSerGlyVal 710
Db 2350 TCCACTGTTCTGCT-----CAGGAGGATGAAGATATTGAAGCTCC----- 2391
Qy 711 SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db 2392 -----AATGAAGAGAAAAATCCAGAGACAGCGGAA----- 2421
Qy 731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
Db 2421 ----- 2421
Qy 751 IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProPro 770
Db 2422 -----GTTGAAGCTGTCAAG----- 2436
Qy 771 AlaThrLeuGlyAlaAspGlyProProGlyProProThrProProArgArgThrSer 790
Db 2437 -----CCAGCGAGGACAGTCTCTGAAATGTACTTCT 2469

QY	791	ArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProPro	810
Db	2470	CGAGGAAACACAGAACTCGGGTTGAGTTGAG-----CCACACGAGAAACT	2517
QY	811	AlaPro---ProSerProSer---AlaProProValValProLysGluGluLysGlu	828
Db	2518	GCACCCAGTACATCTCCCTCTTAGCAGTTCCAAGTACAAAACACAGCTGAAGATGAAAGT	2577
QY	829	GluGluThrAlaAlaAlaProProVal-----	837
Db	2578	GTGGAGACCCAGGTGAATGACAGCATCACTGCTGTGAGACAGCAGAGATGATGTAGAT	2637
QY	838	-----GluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu	851
Db	2638	CACGAGGACACAGTGTCTGNAGAGGGTTCTGTTGTGATCCCCACCCGCTACCAAA---	2694
QY	852	AlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGlu	871
Db	2695	---GCTGACTCTGTGGACGTTGAAGTCAGGCTGCCAGAAAACCATCATCTAAAGTTGAA	2751
QY	872	GluGlyProAlaLysGlyLysAspAlaGluAlaGluAlaThrAlaGluGlyAlaLeu	891
Db	2752	GGTGATTAATACCAAGAAAGAGACTTGGATAGAGCC-----	2787
QY	892	LysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAla	911
Db	2788	---AGTCAGAAAGGTGGAACCTAGAGATGAAGATTTGGTGGTAGCTCAGCAATAAATGCC	2844
QY	912	-----ProGln-----AspSerAspSerSerAlaThrCysSerAlaAspGluVal	926
Db	2845	CAAAAGCCCGAGCCCGCCAGTCAGCAATGATTCAGTGCACGTCGACGCGTGATGAG---	2901
QY	927	AspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerPro-----ArgProSer	943
Db	2902	---GATGTGATGAGAGCCAGAGCAGAGATGTTTCTATGGACTCAAGCCCTTCA	2958
QY	944	LeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeu	961
Db	2959	CTGTAAACCCCACTGGATCTATACTCGTC---TCATCTCCGTTAAACCAATCCACTCG	3015
QY	962	AspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProlile-----	976
Db	3016	GATCTGCCACAGCTTCAGATCGAGCTGCTGTATCCACCAATGTTATCTCGACCCCA	3075
QY	977	-----GlnValThr	979
Db	3076	TGTAACATACCAATTGGAAACCCAGTCGAGCGGTATGCTCTACCAGCAGACATATAA	3135
QY	980	LysValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaPro	999
Db	3136	GCAATGTGATGAGTCAGCACTCCTCGAG-----	3162
QY	1000	ProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerPro	1019
Db	3163	---GAGCAGCCGCAGACACAGACAGATAGATTGGGAATGTAGAAGTTCTTACAAGTCCA	3219
QY	1020	ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla	1039
Db	3220	TGTGGCACATCCAAGAGTCCA-----AACAGAG--	3249
QY	1040	GlnLysLeuProGlyAspProProCysThrThrSerGlyLeuProPheProValProPro	1059
Db	3250	-----TGG-----	3252
QY	1060	ArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaPro	1079
Db	3253	---GAAGTCCITCAGCTGCTCCACAT-----CAATTGATAAACAATCTCCCT	3297
QY	1080	ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro	1099
Db	3298	GAAAGCGGTTCCGGCTTCGG-----ACAACTGACCACACCGGCCACCGGCC	3342

1100	ProThrIleSerAsnProProProLeuIleSerSerAlaIysHisProSerValLeuGlu	1119	ArgGlnIleGlyValaIleSerGlnGly	1134	MetSerValGlnLeuHis
QY		QY		QY	
3343	CCTCTCATCCCGCTATCCAAAACACACAGTGGCTTCAGAAAAA---CCATCTTTTATA---	3396	ATGGGAGGGCTCCATCTCACAGGGGAACACACGCGCACTTATTGACTTCTCATAAATCAG	3453	ValProTyrSerGluHisAlaIysAlaPro---ValGlyProValThrMetGlyLeuPro
DB		DB		DB	
1120	ArgGlnIleGlyValaIleSerGlnGly	1134	MetSerValGlnLeuHis	1173	CGGCAACAGGAATCTGCCAAATCAGCTACTTTGCCCTACATCAACAGGAGAAATTTCT
QY		QY		DB	
3397	ATGGGAGGGCTCCATCTCACAGGGGAACACACGCGCACTTATTGACTTCTCATAAATCAG	3453	ValProTyrSerGluHisAlaIysAlaPro---ValGlyProValThrMetGlyLeuPro		
DB		DB		QY	
1135	ValProTyrSerGluHisAlaIysAlaPro---ValGlyProValThrMetGlyLeuPro	1153	GCTTCTCATCACTCAAGAACACCCCAAGCGCTCAGTAGGATCTATCTCTCTTTGGACTGCCA		
QY		QY		DB	
3454	GCTTCTCATCACTCAAGAACACCCCAAGCGCTCAGTAGGATCTATCTCTCTTTGGACTGCCA	3513	LeuProMetAspProIlystIysLeuAlaProPheSerGlyValIysGlnGluGlnLeuSer		
DB		DB		QY	
1154	LeuProMetAspProIlystIysLeuAlaProPheSerGlyValIysGlnGluGlnLeuSer	1173	CGGCAACAGGAATCTGCCAAATCAGCTACTTTGCCCTACATCAACAGGAGAAATTTCT		
QY		QY		DB	
3514	CGGCAACAGGAATCTGCCAAATCAGCTACTTTGCCCTACATCAACAGGAGAAATTTCT	3573	ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSer		
DB		DB		QY	
1174	ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSer	1193	ValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrIysGlyIlePro		
QY		QY		DB	
3574	CCCGGAAGCCAAATCTCAACCTGAGGGCTCTGTTGGTC---AGGGCCCAACATGAAGT	3633	GTAGTCAGAGGTACCGCA---GGAGCGCATACAGAAGGAAGTATTAACTCGGGGAATCCA		
DB		DB		QY	
1194	ValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrIysGlyIlePro	1213	SerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThr		
QY		QY		DB	
3631	GTAGTCAGAGGTACCGCA---GGAGCGCATACAGAAGGAAGTATTAACTCGGGGAATCCA	3687	ACCAGCAAAATTTCACTGGAGAGCATTCATCCCTACGGGGCTCTATCACTCAGGGCAC		
DB		DB		QY	
1214	SerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThr	1233	AspValLeuTyrIysGlyThrIleThrArg		
QY		QY		DB	
3688	ACCAGCAAAATTTCACTGGAGAGCATTCATCCCTACGGGGCTCTATCACTCAGGGCAC	3747	CGGGCTCTGCCCAGACTGGCATACACAGAGGCTTTGGTGAAGGGTTCATTTTCGAGA		
DB		DB		QY	
1234	ProAla-----AspValLeuTyrIysGlyThrIleThrArg	1245	IleIleGlyGluAspSerProSerArgLeuAspArgGlyValGluAspSerLeuProIys		
QY		QY		DB	
3748	CGGGCTCTGCCCAGACTGGCATACACAGAGGCTTTGGTGAAGGGTTCATTTTCGAGA	3807	ATGCCCATTTGAACACAGCTCTCT-----GAGAAAGGCAGAGAGGAAGCTGCATCCAAA		
DB		DB		QY	
1246	IleIleGlyGluAspSerProSerArgLeuAspArgGlyValGluAspSerLeuProIys	1265	GlyHisValIleTyrGluGlyIysIysGlyHisValLeuSerTyrGluGlyGlyMetSer		
QY		QY		DB	
3808	ATGCCCATTTGAACACAGCTCTCT-----GAGAAAGGCAGAGAGGAAGCTGCATCCAAA	3861	GGCCATGTTATTATGAAGCAAAAGCTGGACATATCTTCTCATATGATATATTAAGAAT		
DB		DB		QY	
1266	GlyHisValIleTyrGluGlyIysIysGlyHisValLeuSerTyrGluGlyGlyMetSer	1285	ValThrGlnCysSerIysGluAspGlyArgSerSerGlyProProHisGluThrAla		
QY		QY		DB	
3862	GGCCATGTTATTATGAAGCAAAAGCTGGACATATCTTCTCATATGATATATTAAGAAT	3921	CGAGHAGGACATAGAGTCCAAAGACAGCTCATGAATCAGT		
DB		DB		QY	
1286	ValThrGlnCysSerIysGluAspGlyArgSerSerGlyProProHisGluThrAla	1305	AlaProIysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer		
QY		QY		DB	
3922	GCC-----CGAGHAGGACATAGAGTCCAAAGACAGCTCATGAATCAGT	3966	TTA---AAGAGAAGCTATGAATCAGTCCGAAGGAATATAAAGCAAGGGATGTCAATCAGG		
DB		DB		QY	
1306	AlaProIysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer	1323	SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg		
QY		QY		DB	
3967	TTA---AAGAGAAGCTATGAATCAGTCCGAAGGAATATAAAGCAAGGGATGTCAATCAGG	4023	GAGTCTCTGTATCAGCACCGTTAGAGGGCTGATATGCCAGCATTCACC-----AGG		
DB		DB		QY	
1324	SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArg	1339	HisSerProHis---HisLeuLysGluGlnHisIleArgGlySerIleThrGlnGly		
QY		QY		DB	
4024	GAGTCTCTGTATCAGCACCGTTAGAGGGCTGATATGCCAGCATTCACC-----AGG	4077	GGGAGTCTCTATCTCGACCTCAAAGAAAGGACTGTATTGTCTGGCTCCATAATGCAGGGG		
DB		DB		QY	
1340	HisSerProHis---HisLeuLysGluGlnHisIleArgGlySerIleThrGlnGly	1358	IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaIysLeuLeu		
QY		QY		DB	
4078	GGGAGTCTCTATCTCGACCTCAAAGAAAGGACTGTATTGTCTGGCTCCATAATGCAGGGG	4137	ACACCAAGAGCAACCACTGAAGCTTTGAAGATGGCCCTT---AAATATCCCAACCAATT		
DB		DB		QY	
1359	IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaIysLeuLeu	1378	LysArgGlu		

Db 4333 CGAAAACTCCAGAGTGGTCCAGAGCACCGCCGATATTCAGGGTTCATTTCCCGAG 4392
Qy 1454 GlyThrProLeuLeuValThrGlyAlaSerThrGlySerLysLysHisAspVal 1473
Db 4393 GGCACACCAATAAGTTTGAACAAC---TCAGGTCAATCTGCCATCAACAATATGTC 4449
Qy 1474 ArgSerLeuLeuGlySerProGlyArgThrPheProProValHisProLeuAspValMet 1493
Db 4450 AAATCTCTTAATCAGCGGCTAGCAAACTATCCCGTGAATGCTCCGCTGGAAATTTGTG 4509
Qy 1494 AlaAsp---AlaArgAlaLeuGluAlaArgAlaCysTyrGlu-----GluSer 1507
Db 4510 CCAGAGAACAATAAAGTGTAGAACCGGGGAAATATAGAGATGTGAAAGCAGCGAGACC 4569
Qy 1508 LeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaPro 1527
Db 4570 GTCCGTTCCCGCACACGTCAGTGTGAAGTCTGGCCCTCCGTTCTTAGTCCACA--- 4626
Qy 1528 ValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGly 1547
Db 4627 -----CTGCATGAAGCTCCAAAGCACAACTAGCCCTGGGATTTATGATGACACCACT 4680
Qy 1548 Ala-----ProPheAlaGlyHisLeuProArgGlySerProValThrMet 1562
Db 4681 GCACGGAGGACCCCTGTGATTTATCAAAACACCATGTCCAGAGGCTCACCCATGATGAAC 4740
Qy 1563 ArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAsp 1582
Db 4741 AGAATCTCTGATTTACAATCTCT-----CCTAACAGTCTACCAATCATGAA 4788
Qy 1583 ArgLysLeuThrSerThrProArgGlu-----IleAlaLysSerProHisSer 1598
Db 4789 AGGAATCGACACTGACCCCTACCCAGAGGAGAAAGTATCCAGCGAAGTCTCCAGTGCCT 4848
Qy 1599 ThrValProGluHisProHisProIleSerProTyrGluHisLeuLeuArgGlyVal 1618
Db 4849 GGGGTGGACCCCTGCTGTGAGCCAC-----AGTCCGTTTGTATCCCATCCACAGGCGAGC 4902
Qy 1619 SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
Db 4903 ACTGCGAGCGAGGTTTATCGAGCCACCTCCCGCAGCAATGGATTCAC---GCCATGCT 4959
Qy 1639 ArgGlyIleProLeuAsp---AlaAlaAlaTyrTyrLeuProArgHisLeuAlaPro 1657
Db 4960 TTTCACAGGCTTTGGATCTCGAGCGCTGCTTACCTGTTTCAGAGACAGCTTTCACCA 5019
Qy 1658 AsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAla 1677
Db 5020 ACTCCAGGTTACCAAGTCAAGTATCAGCTTTAC----- 5052
Qy 1678 AlaLeuGluAsn---ArgGlnThrIleLeuAsnAspTyrIleThrSerGlnMetHis 1696
Db 5053 GCAATGGAGAACACAGACAGACAATCTTAAATGATTACATTACCTCAACAAGATGCAA 5112
Qy 1697 HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArg 1716
Db 5113 GTGAACCTG-----CGTCCAGATGTGGCCAGAGGACTCTCCCAAGA 5154
Qy 1717 GluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleLeuAspLeuSerGln 1736
Db 5155 GAGCAGCACTGGGTCTCCCATACCCAGCA---ACGAGAGGAATCATTTGACCTGACCAAT 5211
Qy 1737 ValProHisLeuProValLeuValProThrProThrProGlyThrProAlaThrAlaMetAsp 1756
Db 5212 ATGCTTCCA---ACAATTTAGTGGCTCATCCAGGGGGAACAAGCACTCTCCCATGGAC 5268
Qy 1757 ArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg---HisSerSerSer 1775
Db 5269 AGAATCACTTATTTCTGTGTACACAGATTACTTCCCTCCAGCGCGTACACTCTGCT 5328
Qy 1776 ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrThrSerSerGlu 1795

Db 5329 TCCATGCTTCCAGGACACCCAAACACACCTTT-----GCAGCTGCTGCAAGTGTCTGAG 5379
Qy 1796 ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeu 1815
Db 5380 AGGGAACGGGACGGAGCGGAGAGAGCGGGAGCGGCGGACCG-----ATTGCT 5430
Qy 1816 ThrSerThrThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSer 1835
Db 5431 GCAGCTTCTCCGACCTC-----TACCTCGCGCCAGCTCAGAACAG----- 5472
Qy 1836 GlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855
Db 5473 -----CCTGGCCGACCTGGCACTCATGGA 5496
Qy 1856 HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArg 1874
Db 5497 TATGTTCCGCTCCCTTCCCT---TCAGTAAGAAGCTCAGAGACCATGTTGCAACAGAGA 5553
Qy 1875 ProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLys 1894
Db 5554 CCAGTGTGTTTCCAAAGGAACCAATGGAACCAAGTGTATCACACCTTTGGATCCAACCTGCT 5613
Qy 1895 ProThrValLeuArgSerThrSerSerProVal-----ArgProAla 1910
Db 5614 CAGCTACGAATCATGCTGCTGCTGGGGCCCTTCAATAAGCCAAGCCTGCCAGCC 5673
Qy 1911 AlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPro 1930
Db 5674 TCCCGTTACACACTGCTCGGATGCC---CTGGCTGCTTCTGTGGATGCTGCAGCTTCT 5730
Qy 1931 ThrLeuMetGluProValLeuLeuProLys-----GluAlaProArgValAla 1946
Db 5731 GCACCCAGATGATGTGTCCAAACAAAGAGAGTAGTAAGCATGAAGCTGCAGGTTAGAA 5790
Qy 1947 ArgProGluArgProArgAlaAsp----- 1954
Db 5791 GAAATTTGAGAGCAGGTCAGCAGCAGTGTAGTGAACAGCAGCAGTAGAGCAAGAAACC 5850
Qy 1955 -----ThrGlyHisAlaPheLeuAlaLys 1962
Db 5851 CTGGAGGTGGAGNAGAGATCTGTTCACTGTTTATACACTTCTTCAGCTTCTCCAGTGGC 5910
Qy 1963 ProProAlaArgSerGlyLeuGluProAlaSerSer-----ProSerLysGlySer 1979
Db 5911 AAGCCC-----CAGCCTCATCTTCTAGTAGTTTATCTAGGCTGGGAAA 5955
Qy 1980 GluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAla 1999
Db 5956 GATAAAGGCT-----CCTCAAAATCCAGATATGAGGAAGAGCTAAGGACAGAGGG 6009
Qy 2000 LysAsn-----LeuAlaProHisHisAlaSerProAspProProAlaProAlaSer 2017
Db 6010 AAGACTACCATTAATCTGACGTAACCTCATAGAGTGTATCATCACCCGGCAAAATGCTCG 6069
Qy 2018 AlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnLeuGlu 2037
Db 6070 GACAAGATGCGAGGGAACGTGGCTCTCAAAAGTTACAGACTCTTCT----- 6114
Qy 2038 LeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSer 2057
Db 6115 ---AGTAGTTATCTTCTCACAGGTATGAACACCTAGGCTGCTATGAGGTGATTAAGT 6171
Qy 2058 ProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeu 2077
Db 6172 CTGCGCAGCTCCTCGGCCACCCAGGAGAAACTGCAGACCTATCAGCCAGAGGTTGTT 6231
Qy 2078 AspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGly 2097
Db 6232 AAGGCAATCAAGCGGAAATGATCTCTACCAGACAATATGAAGGACCA----- 6279
Qy 2098 GlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSer 2117
Db 6280 -----TTACATCATCTATGACCA-----CAGCAGGAATCACCATCT 6315

Db 3682 TCTGGCTCCATAATGTCAGGGGACACCAAGAGCAACAACACTGAAAGCTTTGAGATGGCCTT 3741
Qy 1372 ArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProSerArg 1391
Db 3742 ---AAATATCCCAACAAATTTAAAGGAAAGT-----CCTCCCATACGA 3783
Qy 1392 AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHis 1411
Db 3784 GCATTGGAAGTGCCATT-----ACCAAGGAAACCA---TAT 3819
Qy 1412 GluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGlu 1431
Db 3820 GATGCG---ATCACACCATCAAGAAATGCGGGCTTCCATTCAITGAGATTCCAAAGCAA 3876
Qy 1432 GluLeu-----ArgHisThrProGluLeuProLeuAlaProArgProLeu 1446
Db 3877 GATATTTTAACTCAGGAAAGTCGGAAGTCCAGAGTGGTCCAGAGCACACGCGCGATA 3936
Qy 1447 LysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThr 1466
Db 3937 ATTGAGGGTTCCATTTCACGGGCACCAATAAAGTTTGACAACAAC---TCAGGTCAA 3993
Qy 1467 GlySerLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProPro 1486
Db 3994 TCTGCCATCAACACAATGTCAATTCCTTAATCACGGGCGCTAGCAAACTATCCCGTGGGA 4053
Qy 1487 ValHisProLeuAspValMetAlaAsp---AlaArgAlaLeuGluArgAlaCysTyrGlu 1505
Db 4054 ATGCTCTCGTGGAAATGTGCCAGAGNACATAAAGTGGTAGAACGGGGAATATGAG 4113
Qy 1506 -----GluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGly 1520
Db 4114 GATGTGAAGCAGCGAGACCGTGGTTCGCGGCACACGTCAAGTGAAGCTCTGGCCCC 4173
Qy 1521 SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPro 1540
Db 4174 TCCGTTCTTAGTCCACA-----CTGCATGAAGCTCCCAAGACCAACTCAGGCCCT 4224
Qy 1541 LeuThrTyrGluAspHisGlyAla-----ProPheAlaGlyHisLeuPro 1555
Db 4225 GGGATTATGATGACACCATGTCAGGAGGACCCTGTGAGTTATCAAAACACCATGTCC 4284
Qy 1556 ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSer 1575
Db 4285 AGAGGCTCACCATGATGAACAGA-----ACTTCTGATGTTCT 4323
Qy 1576 SerSerLysAla---SerGlnAspArgLysLeuThrSerThrProArgGlu----- 1591
Db 4324 TCTAACAAAGTCTACCAATCATGAAGGAAATCGACACTGCCCTTACCCAGAGGAAAGT 4383
Qy 1592 ---IleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPro 1610
Db 4384 ATCCAGGAACTCTCCAGTCCCTGGGTGGACCCCTGTCGTGAGCCAC-----AGTCCG 4437
Qy 1611 TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeu 1630
Db 4438 TTTGATCCCATCACAGAGGCAGCACCTGACGGCGAGGTTTATCGAGGCCACTGCCACG 4497
Qy 1631 AlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyr 1650
Db 4498 CACTTGGATCCA---GCCATCCCTTTTCCAGAGGCTTTGGATCTCGCGGCTGTACTCTG 4554
Qy 1651 LeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIle 1670
Db 4555 TTTCCAGAGACAGCTTTTCCCAACTCCAGGTACCCAGTCAAGTCAAGTTCAGCTTTAC 4608
Qy 1671 ArgGlyTyrProAspThrAlaAlaLeuGlnAsn---ArgGlnThrIleIleAsnAspTyr 1689
Db 4609 -----GCAATGGAGAACAACAAGACACACACATCTTAATGATTAC 4647
Qy 1690 IleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMet 1709
Db 4648 ATTACCTCACAAACAGATGCAAGTGAAGTCTTG-----CGTCCAGATGTG 4689

Qy 1710 LeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArg 1729
Db 4690 GCCAGAGGACTCTCCCAAGAGAGCAGCAGCTGGTCTCCCATACCCAGCA---ACGAGA 4746
Qy 1730 GlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGly 1749
Db 4747 GGAATCATTTGACCTGACCAATATGCTCCA---ACAATTTTAGTGGCTCATCCAGGGGA 4803
Qy 1750 ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSer 1769
Db 4804 ACAAGCACTCTCCCATGGACAGATCACTTATATTCTGGTACACAGATTACTTTCCCT 4863
Qy 1770 SerArg---HisSerSerSerProLeuSerProGlyThrProHisLeuThrLysPro 1788
Db 4864 CCCAGGCGGTACAACCTCTGCTCCATGCTCTCCAGGACACCCCAACACACCTT----- 4914
Qy 1789 ThrThrThrSerSerGluArgGluArgAspArgAspArgGluArgAspArgAspArg 1808
Db 4915 GCAGCTGCTCAAGTGTCTGAGAGGAAACGGGACGGGAGCGGAGAGCGGAGCGG 4974
Qy 1809 GluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaProIleTyrArg 1828
Db 4975 GAACGG-----ATTGTGACAGCTTCTCCACCTC-----TACTGCGG 5013
Qy 1829 ProGlyThrGluGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlySer 1848
Db 5014 CCAGCTCAGAACAG-----CCT 5031
Qy 1849 SerArgProAlaSerHisHisAlaHisGlnHisSerProIleSerProArgThrGln 1868
Db 5032 GGCGCAGCTGGCAGTCAATGATATGTTGCTCCCTTCCCT---TCAGTAAGAACTCAG 5088
Qy 1869 AspAla---LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIle 1887
Db 5089 GAGACCATGTTGCAACAGACCCAGTGTTCACAGAACCAATGGAACACAGTGTATC 5148
Qy 1888 ThrAlaValGlnProSerLysProThrValLeuArgSerThrSerThrSerProVal 1907
Db 5149 ACACCTTTGGATCCACTGCTCAGTACGAATCATGCCACTGCTGCTGGGGCCCTTCA 5208
Qy 1908 -----ArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGly 1923
Db 5209 ATAAGCCAAAGCCTGCCAGCCTCCCGTTACAACTCTCGGATGCC---CTGGCTGCT 5265
Qy 1924 ThrLeuAspGlyValTyrProThrLeuMetGluProValLeuProLys----- 1940
Db 5266 CTTGTGATGCTGCAGCTTCTGCACCCAGATGATGTGTCCAAACAAAGAGAGTAAAG 5325
Qy 1941 ---GluAlaProArgValAlaArgProGluArgProArgAlaAsp----- 1954
Db 5326 CATGAAGCTCCAGGTTAGAAGAAATTTGAGAGAGGAGTTCAGCAGCAGTGTAGTGAACAG 5385
Qy 1955 -----Thr 1955
Db 5386 CAGCAGCTAGACAGCAAAACCTCGAGGTGGAGAGAGATCTGTTCAGTGTTTATACAT 5445
Qy 1956 GlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSer--- 1974
Db 5446 TCTTACGCTTCCAAAGTGGCAAGCCC-----CAGCCTCATCTTTCAGTA 5490
Qy 1975 -----ProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAla 1992
Db 5491 GTTTATTCTCAGGCTGGGAAGATAAAGGCT-----CCTCCAAATCCAGATATGAG 5544
Qy 1993 ThrIleAlaArgThrProAlaLysAsn-----LeuAlaProHisHisAlaSerProAsp 2010
Db 5545 GAAGAGCTAAGGACAGAGGAGAACTACCATTAATCTGACGTAACCTTTCATAGACGTGATC 5604
Qy 2011 ProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysPro 2030
Db 5605 ATCACCAGGCAATGCTCTCGACAGAGATGCGAGGAAACGTGGCTCTCTCAAAAGTTTACAG 5664

QY 2031 PheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerPro 2050
DB 5665 TCCTTCT-----AGTAGCTTATCTCTCACAGGTATGAACACCTAGC 5706
QY 2051 GluGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeu 2070
DB 5707 GATGCTATTGAGTGATAAGTCTGCCAGCTCACTTGGCCGCCAGGAGAACTGCAG 5766
QY 2071 ProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGln 2090
DB 5767 ACCTATCAGCAGGAGGTGTTAAAGCAATCAACGCGAAATGATCTTACAGACAATAT 5826
QY 2091 ProGlyProValLysLeuGlyGluAlaHisLeuProHisLeuArgProLeuPro 2110
DB 5827 GAAGGACCA-----TTACATCACTATCGACCA----- 5853
QY 2111 GluSerGlnProSerSerProLeuLeuGlnThrAlaProGly----- 2125
DB 5854 ---CAGCAGGAATCACCATCTCCCAACACAGCTGCCCTTCTTCCACAGCGAGGGA 5910
QY 2126 -----ValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluVal 2142
DB 5911 ATGGGGCAAGTGGCCAGGCCATCGGCTGATCACACTTGTCTGATCACTGTCAAAAT 5970
QY 2143 IleThrGlnAspTyrThrArgHisHis-----ProGlnGlnLeuSerAla 2157
DB 5971 ATCACACAAGATTGTGTAGAAATCAAGTTTCTCGCAGACTCCCGACGAG----- 6021
QY 2158 ProLeuProAlaProLeuTyrSerPheProGlyAla-----SerCysProValLeuAsp 2175
DB 6022 CCTCTACTTCTATCTCCAGAACTCACCTTCTGTGTGTATCTACACCTGTG----- 6075
QY 2176 LeuArgArgProProSerAspLeuTyrLeuProProAsp-----His 2190
DB 6076 ---AGGACTAAACATCAAAACCGTTACAGCCCAAGATCCAGGCTCAGTCTGTCCATCAT 6132
QY 2191 GlyAlaPro---AlaArgGlySerProHis-----SerGluGlyGlyLys 2204
DB 6133 CAAAGACAGGTTCAAGGGTCTCTCCAGAAATCTTGGACAAATCCAGGGGAGTAGG 6192
QY 2205 ArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGlyLysGluPro 2224
DB 6193 CCTGGAATAATCCCCAGAGAGAGTACAGTCT-----TCTTGGAGCCCTACGAGCCC 6243
QY 2225 ValSerProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeu 2244
DB 6244 ATCTCCCCACCCAG-----GTTCCGGTGTGTGATGAGAAACAGGACAGCTTGTCTGCTC 6297
QY 2245 LeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsn 2264
DB 6298 TTGTCTCAGAGGGCGCAGAGCTGCAGACAGAGGAATGATGCCCTCACCGGAGT 6357
QY 2265 ThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLys 2284
DB 6358 ATAAGCTACTTGTCTTCTTCTTCCCAAGCTT---GAAATATACATCACCATGGTTAAA 6414
QY 2285 SerLysLysGlnGluIleAsnLysLeuAsnThrHisAsnArgAsnGluProGluTyr 2304
DB 6415 TCAAAGAGCAGGAGATTTTTCGTAAGTTGAATCTCTCGTGGAGTGTACTGTGATG 6474
QY 2305 AsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyLysLeu 2324
DB 6475 GCAGCTGCTCAGCCAGGAGCTGAGATCTTTAATCTGCCAGCAGTACTACGTCAGGCTCA 6534
QY 2325 MetThrTyrArgSerGlnAlaValGlnGluHisIleSerThrAsnMetGlyLeuGluAla 2344
DB 6535 GTTAGCTCTAGAGCCATCTTTTGTCTGATCCTGCCAGT---AATCTTGGCTGGAAGAC 6591
QY 2345 IleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlu----- 2359
DB 6592 ATTATCAGGAGGCTCTCATGGGAGCTTTTGATGACAAAGTTGGAGATCATGGAGTTGTC 6651
QY 2360 ---SerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerLeuPro 2378

DB 6652 ATGTCCCGAGCTATGGAGTA-----GTGCTGTGTACTGCCAACACCTCAGTT--- 6699
QY 2379 AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGly 2398
DB 6700 -----GTGACCAGTGTGTGAGACACGAGAGAGGAAGAGGGGACCCCATCACCTCAT 6747
QY 2399 GlyGly---GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerPro 2417
DB 6748 TCAGGAGTTTGCACCAAGCTGATCAGCAAGTCAACAGCAGGAGAAATCTAAGTCTCCT 6807
QY 2418 AlaPro-----GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSer 2435
DB 6808 ATACCTGGCGGAGGCTACTTAGGAACGGAACGGCCCTCTTCACTCTCTCTGTACATTC 6867
QY 2436 GluGlyAspCysAsnArgThrProLeuThrAsnArgValTrpGluAspArgProSer 2455
DB 6868 GAAGGGGATTTACCATAGGAGAGCCCA-----GGGTGGCCCTGGGAGAGACAGCCCTCT 6921
QY 2456 SerAlaGlySerThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyVal 2475
DB 6922 TCACAGGCTCACTCAGTTTCTTATACCTCTCACTATGCGATG-----CTC 6972
QY 2476 MetAlaSerProProProGlyLeuProAlaGlySerGly---ProLeuAlaGlyPro 2494
DB 6973 AGCAGTACTCCACCAACACGATTCATGTGTCTCTCTCGGTGAACCAAGCAGCTCCT 7032
QY 2495 His-----HisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGlu 2511
DB 7033 CACCAACAGAACAGGATCTGGGAGCGAGAGCTGCCCTCTCTCAGCACAGTACGAG 7092
QY 2512 ThrLeuSerAspSerGlu 2517
DB 7093 ACCCTGTGGATAGTGTAT 7110

RESULT 13
AF495886
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

AF495886 8959 bp mRNA linear VRT 25-NOV-2002
Xenopus laevis nuclear receptor corepressor mRNA, complete cds.
AF495886
AF495886.1 GI:20147786
Xenopus laevis (African clawed frog)
Xenopus laevis
Xenopus laevis
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 8959)
Sachs, L.M., Jones, P.L., Havis, E., Rouse, N., Demeneix, B.A. and
Shi, Y.B.
Nuclear Receptor Corepressor Recruitment by Unliganded Thyroid
Hormone Receptor in Gene Repression during Xenopus laevis
Development
Mol. Cell. Biol. 22 (24), 8527-8538 (2002)
12446772
2 (bases 1 to 8959)
Sachs, L.M., Jones, P.L. and Shi, Y.-B.
Direct Submission
Submitted (27-MAR-2002) Lab. Physiology, UMR 8572 CNRS, Museum
National d'Histoire Naturelle, 7 rue Cuvier, Paris 75231, France
Location/Qualifiers
1. .8959
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
499. .7995
/codon_start=1
/product="nuclear receptor corepressor"
/protein_id="AA012658.1"
/db_xref="GI:20147787"
/translation="MSSGGYPPNQAFSTEQGRYSSHPVOYTFPSRHHQOQYSEYR
TALEASQLTIQQQLRRRPSLSEFHPVSDRPPDRRSYDQYQTHSISQNEHALESKR

PRLDQVSDSHVRLNPAMVLLVPTIOBQVRVVOSEVKEQGLSSKHSSSSPLSGQPGDD
PDASPSKLSKEELISDMRVDREIAKVEQQLLKLKKQQLLEEAAPKEPEKVPSP
PEQKHSVQVQIYIDENRKAEEAKFLEGLPKVLEPLYNQPSDTYHNIENKTV
MKKLIILFKFRNARLAKLEONIQRVDQLMEAEKVKVDRLENNPRKAKESKREY
EKOPETRKOREQOERFORVQORCTGMSATARSEHEISETIDGLSPKAKENKQMRLO
VLIIPMMEDAEORVKVINTNGLMEDPKVYKORQFNVMTDHEKELPKKPVHRPKN
FGLIASYLERNSDCVLYIYIKKENLKSILVRNTPKGRNQOQITRPSQBEKEI
EKVEEKADNRDKKEDEEEKEELEREGAKIDDAVEDGDDEKQDSMPGRKT
ANNQGRKRVTRSMASAAAANAATVATTAVTTTATVAPVPVAPPEPTPP
PPOQSILVDGRNNGALAKMVGSKEQCKNFYNYKRRNLNLOLQHKQKSSRRR
EEDVSDCSDTASTVSAQEDENEAENPEDESEGENSDTESAPSPAEABL
GDSDVDRTSVSMNEAPEODAAKSPADSPSPISVENTPEPTQYAEIKVEIESTE
NEQTMVEBERSQSALSNLQVTKAETETEAKTETSTTVKVTESKEDDSMER
LMDRAEADMGYAPQNIQAROBESQPDNDSSATSEADEEDGDEPDRMYRQSLRS
LLNQAGFLFKQPMDLQLOHRAAVIPMTSCSPCNLTGTATAASNFMSYGRLYEN
SLLEQORQOELTLNRMGASPGNMWEGKSVYMPTEVXHPFEHLAOMON
VARSVPYRLSPRDVRSQVDMNPSRYCVPPVLPAPHOVITSLSDSARLPVTRPT
RPPPLIPSSKTSATSDKSPFITGSGISQGTFTYLASIQYSQBTLPKSPVGSISL
GLPQHESTKICGVYIKQEEFPSPRQSSQPEGLLVRAQHEGVVRSMAAIOEGSIAR
GTPATKVPLEAVSTLRSITQGTALQSQGIADVLKSTITRLAEDIDGSPERCRE
TSAGHVIYEGKSHIYSDTIKMRGEGTSPRTAVALKRAYDTMEGNIKOAMSVR
EAAVSGMEGLICTLPGKSPHPELKDORVLSGIMKGTPTTNDNDDGLKYAKQIK
LESPIRPSFGAISKGPVEGVTTIKELGRSHIHEIPQDLVSOESRKTPESSQIMEG
SISQGTPIKVESTSQSAIKHNVKSLITGPSNLGWSVQDMNPENLKWGERSKHEDT
KSDAIRSRHTSVVSGPSVLIRSTLHEASKSQSPGIYEDNNARHTPVNYPMSRST
PMARSABEGCLTPSKSSHERKSTLPTQRESIVKSPVGVDPSPSAHSPFDTHLSAP
PGDYRAHLPHLDLPALQFHRPLDPAAYLFLQRLSPTGCPYQOLYAMENTROTPI
LNDYITSQOMQVNLPOVARGLSPRDQGLGIPYARGIIDLNTMPAILLVHPHGTS
TPMDRTIYIPGTLAPFPRRYPNAPSPGHPAHLANSVSAEREREEREREDRE
KEQBERSDERERDLAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA
SPVSAVESIMQORPSIFLPTNGKSVITPLDAAQFRIMPPTPGAASITQIPASRST
AADALAALDAATAPOMEVAKPKESKNDSEENSRNALLEQQOQIDCERRVMS
PYTSSPSSSQSSQSSAVYSEAGKERTAKTSRYVEELMRGKTTITAAFNIDVIT
TOOIASDKGRNSQSSDASSSHSRYEAPRETI EVI SPANSPVQEKESYPOEMPK
SSOTETGEGKVDGPNRYQOQESPQOTIPGHVPOTHRLITLADHICHIITQDA
RNQPNVQLOQPASTQSTNPSTAVRTKASRSPESQVPMPSRPSASRSPENSI
PDKPRGPGSPORGLISEPEYIISPQAPMLHTKQDNMLLSORQPPRNDSP
GNLYLSPFTKLEPSPMYMKKEIFRLKNSGGSDMAAAQPOTEIFNLPAVTT
SGAISRSHGSPADPNLGLIEDIRKALMGNFODKSDHGVLPQVPOGNOGTPNSA
RREANPSPNGSGTHQKLIISKYGRKTSPIGSGOTYLGAEPPSSVSVSEGDYR
QASNAWEDRPSGTSGTQFPYPLNTMGLNLSPPSSMSCTPSTNTQTSAAQQRIWER
EPAPLSEQYETIUSDSB"

ORIGIN

Alignment Scores:

Pred. No.:	3,34e-73	Length:	8959
Score:	4009.50	Matches:	1076
Percent Similarity:	52.41%	Conservative:	348
Best Local Similarity:	39.60%	Mismatches:	839
Query Match:	30.34%	Indels:	455
DB:	5	Gaps:	104

US-09-522-753-5 (1-2517) x AF495886 (1-8959)

Qy	18	ArgTyrProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspVal	37
Db	553	CGATATTCTTCTCATCTGTACATACATCTTCCCGAGCAGCCGACATCAGCAGGAATAT	612
Qy	38	GlyLeuLeuGluTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySer	57
Db	613	TCTGTTTCTGAGTATCCG-----ACTGCTCACTTGGAGAGCATCCAG	654
Qy	58	IleIleGlnProGln-----ArgArgProSerLeuLeuSerGluPheGlnProGly	75
Db	655	CTGATACAGCAGCAGCAGCTTCGAGGGCGTCTTATTTGTTCAGAAATTCACCCAGTG	714
Qy	76	AsnGluArgSerGlnGluLeuHis-----LeuArgProGluSerHisSerTyrLeuPro	93
Db	715	TCCGACAGACCCAGGACAGAGATCAGGCTATGATCAACAATACCATTC-----	765
Qy	94	GluLeuGlySerGluMetGluPheIleGluSerIlySerArgProArgLeuGluLeu	113
Db	766	---ATTTCCTCAATGATGATGAGCCCTGGAATCCAGAGCGCCCGATGATGACACTGCGTT	1899

Qy	114	ProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGlu	133
Db	823	TCGGATTCTCATTACAGACTCAACCTGCAATGGTCTTTTGGTACCACCATCCAGGAG	882
Qy	134	-----AspLeuThrIlyAspArgSerLeuThrGlyLeuGluGluPro	147
Db	883	GGTTCGCGAGTGCAGAAAGTGAAGTTTAAAGAGGAGCAGGGTTGTCCGCAAAATGAGTCT	942
Qy	148	ValSerProProSerProProHisThrAspProGluLeuLeuValProProArgLeu	167
Db	943	TCITTCGTCACTCTTTCTGGGCAACCTGGAGATGATCAGGATGCTTCTCCATCCGAACTT	1002
Qy	168	SerIlyGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGlu	187
Db	1003	TCGAAAGAGAACTAATAACAAAGTATGACCGTGTGACCGTGAAGTTGCTTAAAGTTGAA	1062
Qy	188	GlnGlnIleSerIlyLeuLeuIlyLeuGlnGlnLeuGluGluAlaAlaIlyAspPro	207
Db	1063	CAGCAAAATATTAAAGCTTGAAGAAAAAAGCAGCAACAGCTTGAAGAGAGAGCAGCGAACT	1122
Qy	208	ProGluProGluIlyProValSerProProIleGluSerIlyHisAspSerLeuVal	227
Db	1123	CCAGAGCCAGAGAGCCGCTCTCCCTCCACCTGTGTGAAACAGAGCATCGCAGTGTAGTC	1182
Qy	228	GlnIleIleTyrAspGluAsnArgIlyLeuAlaGluAlaAlaHisArgIleLeuGluGly	247
Db	1183	CAGATTATTTATGATGAGAACCGGAAAAAGCAGAGAGAGCAGCAACAGCTTCTTAGAGGT	1242
Qy	248	LeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHis	267
Db	1243	CTAGGTCAAAAGTAGAATCGCTCTTTATTAACAGCCCTCAGACACAAAAAGTTTATCAT	1302
Qy	268	GluAsnIleIleValIleAsnGlnAlaMetArgIlyLeuIleLeuTyrPheLeuArgArg	287
Db	1303	GAAACATCAAGCAATCAGGTGATGAGGAAAAAGCTTATCTGTTCTTTTAAAGAGACGA	1362
Qy	288	AsnHisAlaArgIlyGlnTyrIlyGlnIlyPheCysGlnArgTyrAspGlnLeuMetGlu	307
Db	1363	AATCATCTAGGAACTAAGGAAACAGAACATTTGTTCAGCGCTATGATCAGCTGATGAG	1422
Qy	308	AlaLeuGluIlyIlyValGluArgIleGluAsnAsnProArgArgAlaIlyGluSer	327
Db	1423	GCCTGGGAGAGAAAGTTGACAGATAGAAAAAATCTCTCGAAGAAAAAGCAGAGAGC	1482
Qy	328	IlyValArgGluTyrTyrGluIlyGlnPheProGluIleArgIlyGlnArgGluGln	347
Db	1483	AAACAGAGAGAAATATTATGAGAGCAATTTCTCGAAATCCGGAAGCAGAGAGAGCAG	1542
Qy	348	GluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArg	367
Db	1543	GAGGTTTTTCAA--AGGGTTGGCAAGAGAGAACTGGAATGTCAGCAACCATAGCCAGG	1599
Qy	368	SerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluIly	387
Db	1600	AGTGAGCATGAAATATCTGAAATAATGACGGCTTTTCAAGAACAGAGAAATAACCAAAAG	1659
Qy	388	GlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnArgIle	407
Db	1660	CAGATGGCAACATCTCAGTCATTTCCATCCCATGATGTTTGTATGACAGAGAAAGAGGTG	1719
Qy	408	IlyPheIleAsnMetAsnGlyLeuMetAlaAspProMetIlyValTyrIlyAspArgGln	427
Db	1720	AAATTATTATACGAATGTTTGTATGAGAGATCCCATGAGGTGTACAGGATCGCAG	1779
Qy	428	ValMetAsnMetTrpSerGluGlnGluIlyGluThrPheArgGluIlyPheMetGlnHis	447
Db	1780	TTTATGAATGTATGAGTATCCAGCAGAGAGAGATTTTAAAGAAAAAATTTGTACGGCAC	1839
Qy	448	ProIlyAsnPheGlyLeuIleAlaSerPheLeuGluArgIlyThrValAlaGluCysVal	467
Db	1840	CCTAAAAAATTTGAGTATGCTGCTCGTCTGTGAAAGAAAGAAATGTTTTCAGACTGCGTT	1899

	Qy	468	LeuTyrTyrrLysThrIleThrylSerLysAsnGluAsnTyrlSerLysValArgArgSerTyrr	487
	Dd	1900	TTATAATTATTTAACTAAGAAAAATGAATAATTCGAAATCCCTTGCGAAGAACAATAT	1959
	Qy	488	--ArgArgArgglylysserGlnglnGlnglnGlnglnGlnglnGlnglnGlnGln	506
	Dd	1960	CCTTAAGAGGAGACGCACCAAGCACAGATAACGCCCTTCCCAGAAGAAAAGGAA	2019
	Qy	507	GlnGlnGlnProMetProArqSsrSerSerglnGluGluLylysAspGluLylysGlu	526
	Dd	2020	ATTGAAAAAGTAGAGGAGAAAGCAGATAGAAATGACAAAAAGGAGNATGAAGAGG	2079
	Qy	527	LysGluAlaGluLylysGluGluLylysProGluValgluAsnAspLylysGluAspLeu	546
	Dd	2080	GAAGAAGAAAGAAAGAGGAGAAAGAGGAGGTAGGAGGGCGCCAAGACAAAT---	2136
	Qy	547	LysGluLysthraSpasphrSerglyGluAspAsnAspGluLylysGluAlaValAlaSer	566
	Dd	2137	-----GATGCTGTACTGAAGATGGGGATGANAAGATCAATCATGCCCCA	2181
	Qy	567	LysGlyVarglysthrAlaAsnSerglnGlynArgArgLyyslVargilethrArgSerMet	586
	Dd	2182	AGGGTCGTAAACTGCCAATACCAGGCGACGAAAGSGTCAGTTACCATCCATG	2241
	Qy	587	AlaAsnGluAlaAsnSergluAlaIleThrProGlnGlnSerLa-----	602
	Dd	2242	GCAAGTGAAGCTGCTGGCGCAATGCTGTATACACGGCAAATCTGCTCCAGTCACCC	2301
	Qy	603	-----GluLeuAlaSerMetGluLeuAsnGluSerSeraArgTrpThrGlu	617
	Dd	2302	ACTAGCACTGCCACCACTGTTCACCTGTTCCTGTGGC-ACCACCTCCAGAGAACCA	2360
	Qy	618	GluGluMetGluThrAlaLys-LysGlyLeuLeuGluHisGlyVargAsnTrpSerAlail	637
	Dd	2361	TCCTCTCCACCACCAAGAGCAAGTCTGGTCGATCATGTTCGNACTGGGGAGCAAT	2420
	Qy	637	eAlaArgMetValglySerLysthrValserGlnCySLysAsnPheTyrrPheaNtyrLy	657
	Dd	2421	TGCCAAGATGGTGGATCAAAAGCGAATCCCAAGTGTAAAAATTTCTACTTTAATACA	2480
	Qy	657	sLysArgGlnAsnLeuaspGluIleLeuGlnGlnHisLylysLeuLyswetGluLysluar	677
	Dd	2481	GAGACGCACAATCTAGACAATTTGCTTTCAGCAGCACAAACAGAAAGTCATCCCGAGGCC	2540
	Qy	677	gaenAlaArgLylysLylysAlaProAlaAlaAlaSerGluGluAlaAlaPhePr	697
	Dd	2541	TCGAGAGGACGCATGTCTCAATGTGACAGTATTGCTTCTACTGTATCTGCA----	2595
	Qy	697	oProValValgluaspGluMetGluAlaSerGlyValserGlyAsnGluGluMe	717
	Dd	2596	-----CAGGAAGATGATGAAATGAAGCTCT-----AATCAGGAGGAAAA	2636
	Qy	717	tValGluGluAlaGluAlaleuHisAlaserGlyAsnGluValproArggylGluCySe	737
	Dd	2637	CCCTGGAAGACAGTGAAGGT-----	2655
	Qy	737	xGlyProAlathrValAsnAsnSerSerAspThrGluSerlleProSerProHistrGl	757
	Dd	2656	-----GCAGAAACAGTTTCAGACACTGAAGTGCCTCCCTCTCTCT--	2694
	Qy	757	uAlaAlalySaspThrGlyGlnAsnGlyProLysProAlathrLeuGlyAlaAspGl	777
	Dd	2695	-----TCCCTGTCTGAAGCAGCCAGATTAGCGCATGAC--	2727
	Qy	777	yProProGlyProProThrProProArqArgThrSerArgAlaProilleGluProTh	797
	Dd	2728	-----TCTGTGTATAGAACCACTTCTTCGTATTCCTATCGGAAGTCC	2768
	Qy	797	rProAlaserGluAlathrGlyAlaProThrProProAlaproSerProSerAl	817
	Dd	2769	TCCAGAACAGATGCTGTACCAACCTGCT-----TCAGATTTCATC	2810
	Qy	817	aProProValValProLylysGluLyseGluGluThrAlalaAlaPlaproVa	837

[illegible]

Db 3765 TCCTCCCCCCTCTGATCCCGTCTCCAAACTTCAGCAACTTCTTCTGATAGCCATC 3824
Qy 1116 rValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSer-----ValGI 1132
Db 3825 TTTCAATACAGG-----GGCTCAATATCACAGGGAACACCTGGCACTTACTTAGCGTC 3878
Qy 1132 nLeuHisValProTyrSerGluHisAlaIleAlaPro---ValGlyProValThrMetGI 1151
Db 3879 ACTCAGCAACCTTACAGCCAGGAACCAATGAAGCCCTCTGTGGGTTCATCTCCCTTGG 3938
Qy 1151 yLeuProLeuProMetAspProLyLeuLeuAlaProPheSerGlyValLyGlnGluGI 1171
Db 3939 CCTACCAAGGCAACATGAATCTACAAATTTGGCTGTGTGCTACCTACATCAAGCAAGAGA 3998
Qy 1171 nLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGI 1191
Db 3999 GTTTCCTTCTCGAGGCCAGAGTTCTCAGCCTGAAGACTACTAGTC---AGAGCAGCAGCA 4055
Qy 1191 uAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLyseGI 1211
Db 4056 TGAAGGTGTTGTACAGAGC---TCCATGGCTGCAATTCAGGAGGGGAGCATCGCTAGAGG 4112
Qy 1211 yIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHi 1231
Db 4113 CACTCCAGCAACCAAGTTCCATTGGAAGCAGTTTCCACCTTGTGGGTCCATTACTCA 4172
Qy 1231 sGlyThrPro-----AlaAspValLeuTyrLyGlyThrII 1243
Db 4173 GGGTACACCACTCTTTTACAGTCTGGTATTGCTGCTGATGTTTGTAAAGTCCACCAT 4232
Qy 1243 eThrArgIleIleGlyGluAsp-----SerProSerArgLeuAspArgGlyArgGluAs 1261
Db 4233 CACTAGACTGGCAGCAGAGGATATTGGTAGTCCG-----GAGAGGTGCAGGAGGA 4283
Qy 1261 pSerLeuProLyGlyHisValIleTyrGluGlyLyGlyHisValLeuSerTyrGI 1281
Db 4284 TACCTCTGCTAAAGGGCATGTTATATATGAGGCAAGAGTGGGCACATTTGATCTTATGA 4343
Qy 1281 uGlyGlyMetSerValThrGlnCysSerLyGluAspGlyArgSerSerGlyProPr 1301
Db 4344 C-----ACTATCAAGATATCGGAGAGGAACCAAGGAGCCCAAGCAGCACC 4391
Qy 1301 oHisGluThrAlaAlaProLyAsArgThrTyrAspMetMetGluGlyArgValGlyArgAl 1321
Db 4392 TGAG-----GTGGCTCTGAAGAGAGCATATACCATCGAGGGAATATTAAGCAAGC 4445
Qy 1321 aIleSer-----SerAlaSerIleGluGlyLeuMetGlyArgAlaII 1335
Db 4446 CATGCTGTAAAGAGAGCCGCTGTATCAGGACCTATGGAAGGACTGTATGCTGTACACT 4505
Qy 1335 eProProGluArgHisSerProHis---HisLeuLyGlnGlnHisIleArgGlySe 1354
Db 4506 GCCT-----AAAGGAGTCCACATCCAGAACTCAAGATAGGCAAGTCTTGTCTGGTTC 4559
Qy 1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgGGL 1374
Db 4560 AATCATGAAGGCACTCCGAGAACCAACATCAGACAACTTTGATGATGCCCTT---AAGTA 4616
Qy 1374 uAlaLyLeuLeuLyAsArgGlyThrProProProProProSerArgAspLeuTh 1394
Db 4617 TGAAGGCAATAAACTTGAAGT-----CCGCCTATAAGATCCTTTGA 4661
Qy 1394 rGluAlaTyrLyThrGlnAlaLeuGlyProLeuLyLeuLyProAlaHisGluGlyLe 1414
Db 4662 GGGAGCCATCTCA-----AAAGGCAAACT---TATGAGGT-- 4695
Qy 1414 uValaThrValLyGluAlaGlyArgSerIleHisGluIleProArgGluLeu-- 1433
Db 4696 -GTCAACACGATTAAGAGCTGGGGCGATCTATTATCATGAATTTCCAGACAAGATCTGT 4754
Qy 1434 -----ArgHisThrProGluLeuProLeuAlaProArgProLeuLyGluGI 1449
Db 4755 AAGCCAGAAAGTGCAGAGACTCTCTGA-----AGTTCAAGGCAAAATTAATGGAAGG 4805

Qy 1449 ySerIleThrGlnGlyThrProLeuLyTyrAspThrGlyAlaSerThrThrGlySerLy 1469
Db 4806 TTTCAATTCACAGGCGCACCCATAAATATGAA-----AGCACTTCAGGCGAGTC 4856
Qy 1469 s-----LyHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProVa 1487
Db 4857 TGCTATCAAAACACAATGTGAAGTCTCTGATAACAGGCCCAAGTAACTATCTCTGGGGAT 4916
Qy 1487 lHisProLeuAspValMetAlaAspAlaArgAlaLeu---GluArgAlaCysTyrGlu-- 1505
Db 4917 ACCTCAGATGACAAATATGCCGAAATTTAAATGGGGACGTTCAAACACAGAGA 4976
Qy 1506 -----GluSerLeuLySerArgProGlyThrAlaSerSerSerGlyLyse 1521
Db 4977 CACCAATCATCTCAGCCCATCAGATCAGGCATACGTCCTGTGTGCAGCTCAGGTCCTATC 5036
Qy 1521 rIleAlaArgGlyAlaProValIleValProGluLeuGlyLySerProArgGlnSerProle 1541
Db 5037 TGTCTCCGATCAACC-----TTGCATGAAGCTTCTAAGTCTCAGTTAAGTCTCTGG 5087
Qy 1541 uThrTyrGluAspHisGlyAla-----ProPheAlaGlyHisLeuProArgGlySe 1558
Db 5088 GATTATGAAGACAAATATGCTAGAAGACTCCAGTC-----AATTAACCC--AG 5135
Qy 1558 rProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLy 1578
Db 5136 TCCCATGTCAAGAGCTCACCTATGGCAGCGTCAGCAGAAGTGTCTTACCCCATCCAA 5195
Qy 1578 s---AlaSerGlnAspArgLyLeuThrSerThrProArgGlu-----IleAl 1593
Db 5196 GTCTTCCAGCCATGAAGAAAAAGCACTTTGACTCCCAAGAGAGAGAGCATAGTTGT 5255
Qy 1593 aLySerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHi 1613
Db 5256 GAAGTCACTTCTCTGGTGGACCCCTTCTGCAGCAC-----AGTCCCTTTGACAC 5309
Qy 1613 sLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAs 1633
Db 5310 TCATCTCCGAAGTCTCCACCTGGAGATGTTTATAGACACATTTGGCTCTCCCTACCTAGA 5369
Qy 1633 pProThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaTyrTyrLeuPr 1652
Db 5370 TCCT---GCTCTTCAGATTCCACAGACCATTAGACCCAGCTGCTGCTTACCTGTTCCA 5426
Qy 1652 oArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGI 1672
Db 5427 AAGCAGCTGTACCCACACCTGTTACCTTAGCTAGTACCACTGTAT----- 5475
Qy 1672 yTyrProAspThrAlaAlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyrIleTh 1691
Db 5476 -----GCCATGGAGAGACACAGGACAGCAATCTCTAAATGACTACATAAC 5519
Qy 1691 rSerGlnGlnMetHisHisIleThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuAr 1711
Db 5520 CTCGCAACAGATGCGAGTTAATTTG-----AGGCGGATGTTGCTAG 5561
Qy 1711 gGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyI 1731
Db 5562 AGGACTTTCCTCCAAAGATCAGGAGCTAGGTATTTCATAT---CCTGGAGCCAGAGGAT 5618
Qy 1731 eIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrPr 1751
Db 5619 TATTGACTTACCATAATGCTCCA---GCAATACTGTCCTCATCCAGGTGGCACAAG 5675
Qy 1751 oAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerAr 1771
Db 5676 TACACCAACCCATGACAGGATTTACTTACATTCTCTGGCACAACAACCTGCTTTTCTCCTAG 5735
Qy 1771 q---HisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLySerProThrTh 1790
Db 5736 GCCTTACAACCTGCGCTCATGTCTCCAGGACATCTCTGCACATTTA-----GCAAA 5786


```
QY 874 proAlaLysGlyLysAspAlaGluAlaGluAlaThrAlaGluGlyAlaLeuLysAla 893
Db 135630 CCGGCT-----GGAGCTCAGCCTACT 135610
QY 894 GluLysLysGluGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGln 913
Db 135609 GACCAG-----TATAGTAGATG-GTTAGCGCACAGACTGCCCTCTCCCATGCG 135563
QY 914 AspSerAspSerAla-----ThrCysSerAlaAspGluVal 926
Db 135562 TCATCTAGATGACTGTATACACATCCATTTGGCAGACGTGCAGC----- 135515
QY 927 AspGluAlaGluGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThr 946
Db 135514 -----CAGGAGGGCAGCAA-----GCCTGACCCAGCACTCTTGGT 135479
QY 947 ProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeu 966
Db 135478 CCT-----ATACATCCAGTTGGTGGCATTTCTGGGGGACCGTG 135440
QY 967 LysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArg 986
Db 135439 AAGCAAGATCAGCCCTC----- 135422
QY 987 GluAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsnLeuGlu 1006
Db 135421 TGACACATGGCCCTGTGGTCCCTGCTGCTGGCATGAGCCAGTCCACATACCA 135362
QY 1006 nProGluSer-----AspAlaProGlnGlnProGly-- 1016
Db 135361 AGCAGGTAGCATGGGAACGTAGCAGCACCTCCACTGACCTGTCACGACCCCTCTTG 135302
QY 1017 -----SerSerPro-- 1019
Db 135301 TCAGGGTCTGTTCACAGCTTGCAGTGTGTTTGAGGGTATGGAGATGGCAGACCTCC 135242
QY 1020 -----ArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAl 1037
Db 135241 CTGACTGCCCACTTAATCTTTTCAGAGAGAACCCCATCTTTCGGGTTTCCCAAC 135182
QY 1037 agluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProVa 1057
Db 135181 TGAGGGCCCAAGACTACCGACTGAGCCGCCACGCTGGTCACTCGGCGCTGCCCTTCCCAT 135122
QY 1057 lProProArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTy 1077
Db 135121 CCCTCCACGGAGGTGATCAAGACTTCCCCACACGCGCGTGACCCCTCTGCTCTCTCTA 135062
QY 1077 xAlaProPro----- 1080
Db 135061 CACACCCCGGTGAGTGGGAGCGGACCCAGTTAGGTCTGTGCACTCGGTGGGAGCCAA 135002
QY 1080 ----- 1080
Db 135001 CTCTTTTCAGTGTGCGAGGGCTGCATGGTCCGATCGGATGAGACCTTGGCTGAAGGTGTG 134942
QY 1080 ----- 1080
Db 134941 ACCTGTCACTCCTGTGAAGCTGGGCACTTCTGAGACTCGGGAAAGCTTCCGCACTACC 134882
QY 1080 ----- 1080
Db 134881 TGACAGCCCATAGGCAGCCCTTGGGAGAACTTGAGCTGGAAAGGGGAGGCGCAGCGTGGGA 134822
QY 1080 ----- 1080
Db 134821 GGTATTTTCCACGATGACTAAGAGGATGGTCATTAGAGGGAGCAGTGGGCAGTCTTGAT 134762
QY 1080 ----- 1080
Db 134761 TGACAGATAGTGTGAGAGACTGTTTGGGGGAGACCCACTAGGTTTCTATTCTTGATGG 134702
QY 1080 ----- 1080
Db 134701 TCCTTACCACCACTAAAGTAATCCAGCGGGCCAGTGATTAACAGCATCAGCAGTGG 134642
QY 1080 ----- 1080
Db 134641 CTGGCCCAAGGTAGAGCTGACCCATGGGACAAGTGTACTGAGGGAGCTTGGAGAAACAGGA 134582
QY 1080 ----- 1080
Db 134581 AAGGCATCCAGGAGGAGAAAGAACAGCATAGCATGTGCAAGGGCCAGAGCACTCTCGAG 134522
QY 1080 ----- 1080
Db 134521 TCGGACACTTTGAGSCACCCGAGCATGTGTACAGCATAACTTTTCCACCCAGCAAGCC 134462
QY 1080 ----- 1080
Db 134461 CAGGGATCCCAAGTCACATGTCATCTGGAATCCAGCCCGAGGTGACTCTCTCTGAAGCAT 134402
QY 1080 ----- 1080
Db 134401 AGGAGGGGATAGCGTGTGACTCTGGGAAAGGTCCCTTTGGCCCTCTGTGAGGAGAACAG 134342
QY 1080 ----- 1080
Db 134341 GATCTAGGAGGGAGTGGAAATCGGGGTGCTGGCTTAGAGCTTGGGTGTGGATCGGGG 134282
QY 1080 ----- 1080
Db 134281 TGGGGGTGTGAGCGGTGCAGCTTACAGAAATCCAGGACTGTGGGTTCGGGTGAGTGTG 134222
QY 1080 ----- 1080
Db 134221 CAAGCACTGAAGCAGAGACATGTGGAGGAGCTCCAAAGAGACACTTCATATTTCTTGTG 134162
QY 1081 -GlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProPr 1100
Db 134161 AGGTACCCCGCTGCTCTGGGCTCCACGATAGTGCAGGCGCCGCTCTCTGCCAGCTCCGCC 134102
QY 1100 oThrIleSerAsnProProLeuIleSerSerAlaLysHisProSerValLeuGluAr 1120
Db 134101 C---ATCTCTAACCCCCACCCCTCATCTCTCTGCAAGCATCCCGGCTACTTTGAGAG 134045
QY 1120 gGlnIleGlyAlaIleSer----- 1126
Db 134044 GCAGCTGGGTGCCATCTCCCAAGTGTAGTACCTCTGGGGGTGGGTCTCCCCATCAGAGGA 133985
QY 1126 ----- 1126
Db 133984 GCCTTTTGAAGCCATGGGTGTTTCCAGGGATGGTACTGGGTGGGAGCGGCTGGGCTCA 133925
QY 1126 ----- 1126
Db 133924 TCCCTGGAGTGGGCGAGCCTACGGGGCTACGGGCTACGGTGTGATCTCAGGCTGTGATTCGCGTGC 133865
QY 1126 ----- 1126
Db 133864 CTCAGGAGGTGACTAAACGTCTGGAGGCTCAGTTTCATCTCTGTTTCAGTGAGAGGGTG 133805
QY 1126 ----- 1126
Db 133804 GGGAGCTGCAGCTGTACAGGCTATCGAGTGTACGCTCATAAGCCCTTGGTTCACCTTGC 133745
QY 1126 ----- 1126
Db 133744 ATCCTACCTGCCCTCAGCACACCGGATGGGAAGCAGTCACTGTGTCTTTAGTGGGC 133685
QY 1126 ----- 1126
Db 133684 ACTGGCCTTGGCAGTGTGTTGGCCTGTCTCCATCTCCAGCTGAGTGGGCGGACAGGCC 133625
QY 1126 ----- 1126
```

Db	133624	CATCTCTGACCTGCATATCTCTTAAGTCAGAGTGGGGCTCTCCCTGTGTCTGAAATCTGGT	133565
Qy	1127	-----GlnGlyMetSerValGlnLeuHisValProTyrSerG	1139
Db	133564	TGTGATCTGTCTTCCCTTACCTTAGCAGGGGATGTCAAGTCCAGCTTCGTGTGCTCAGCTCAG	133505
Qy	1139	luHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProL	1159
Db	133504	AGCATGCCAAGGCCCCCATGGGCCCTCTCACCATTGGGGCTCCCTTCGCGTGAGACCTTA	133445
Qy	1159	ysLysLeu-----	1161
Db	133444	AGAGCT--GGGTAGTTCCTCCCAACCTTTGGTTGCTTCAAACTTGGGCTAAGGGGTGGGC	133386
Qy	1162	-----Alap	1163
Db	133385	AGTCCCTGGTGGAGGGCAGCTTGAGACCTCTGTGTATCTCTCACCCGTCCTTAGGCC	133326
Qy	1163	roPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyProProGlnus	1183
Db	133325	CCTTCAGTGGAGTGAAGCAGGAACAGTTGTCCCTCGGGGTCAAGCTGGGCCACCTGAAA	133266
Qy	1183	erLeuGlyValProThrAlaGlnGlnAlaSerValLeuArg-----	1196
Db	133265	GTCTGGGAGTGCCCACTGCTCAGGAGACCTCTGTGCTGAG--AGGTAAGGCTGGCACAGTC	133207
Qy	1196	-----	1196
Db	133206	ACCTTCACTCTAGAGAGCCTGAAAGCCCACTTACAGAGCTGTTCTCATCTTGTAGAC	133147
Qy	1196	-----	1196
Db	133146	ACCAAGAAGACAGTTGGGTGGCTTTCAGACATGGGGTCTACGAGCTGGGTCTTTTGTCT	133087
Qy	1197	-----GlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerT	1215
Db	133086	TTCTAGGAGACAGCACTGGGTTCCGCCACCACTGGAAGCATCACCAGGGCTCCCCAGTA	133027
Qy	1215	hrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHis-----	1231
Db	133026	CCCCGGGTGCAGACGGCCCC-----AGTACAGAGGCTCTATCACCACAGTAGGTGCTG	132973
Qy	1231	-----	1231
Db	132972	GGGGCAGGGTGTACGGGCAGACATGCCAAGTGGCTTGGGATGGATGAGAGGACCTCCT	132913
Qy	1231	-----	1231
Db	132912	GGGAGACTAGAGACAATGGCTGTAATGTTGATCACTGTGCCGTGGGGTCAGGCCTAGG	132853
Qy	1231	-----	1231
Db	132852	GTGATAAAGAGAACAAAGCTCCTGTGTCTCTGGCAGTTAGTGTGGCAGACCAACCAC	132793
Qy	1231	-----	1231
Db	132792	AGATAGGGATAGCCAGGCTCATACAGGCTGTTCAGGAGCCACAGGGCATAGATGTCTG	132733
Qy	1231	-----	1231
Db	132732	GACTGAGAAAGTTACTTGGGCCAGGAAGAGAAAGTGTGACATGTTATTATGGCCGAAG	132673
Qy	1231	-----	1231
Db	132672	AAATAGAGGAGCAGGGGTGGCAGCTGGCTTGGACATATCACCAGTCAAGTTAGGCATTAAGC	132613
Qy	1231	-----	1231
Db	132612	CAGTGGTGGTGCAGACTTAGTACCACCGAGGAGGAAGATGGCGAGTTCCAGGCGA	132553
Qy	1231	-----	1231
Db	132552	ACCTGAAGCTGTAGCTGTAGAGACCTTGTCTTATGAAACAGTATGGTAGTATTCTCGATA	132493
Qy	1231	-----	1231
Db	132492	GCATCTCTGGGAGTGTCAAGTGTAGAGCAGGGGTTGGCAGCAGCATTTGGCACTCGGAA	132433
Qy	1231	-----	1231
Db	132432	GCCAGGATGAGAGACATTGACAACAATCGGAGGGGAGTTACACATGTGTGGCTGTGCTC	132373
Qy	1231	-----	1231
Db	132372	TGTGCATGCTTGGGGGTGGGGAGTGGGTGTGAGAGTCAACCCACAGATCCGG	132313
Qy	1232	-----GlyThrProAlaAspValLeuTyrLysGlyThrIle	1243
Db	132312	GTGCTGCCACTCCACCCCTACAGGGCAGCCCGCAGACGCTCTCTACAAGGGTACCATC	132253
Qy	1244	ThrArgIlelleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeu	1263
Db	132252	AGCAGGATCGTGGGTGAGGACAGCCCAAGTCGCTTGACCGGGCAGCAGAGGACACCCCTG	132193
Qy	1264	ProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGlu-----	1281
Db	132192	CCCAAGGGCCATGTCATCTATAGGGCAAGAAAGCCACGCTCTATCTCTATGAAGGTGAG	132133
Qy	1281	-----	1281
Db	132132	GACAGAGCAGACAACACGAGGCTCTGGCAGTTAGGTGGGTGGCTATGCCGTCTCTCT	132073
Qy	1281	-----	1281
Db	132072	CCCTTTCTTTCGTTACAGGATGTCTTTTCTCAAGACAGACATAGCAAGGCCACTGAGA	132013
Qy	1281	-----	1281
Db	132012	GCAACCACTTGTAGCTGAGGCTTAGCAAGCAGACATAGCAGATAGTACTTCCCTG	131953
Qy	1281	-----	1281
Db	131952	CCAGCTCGCCCTGCTGGGAGTGTGTGGAAACCTTTTGTGATTAGGAGTCATGAAGAC	131893
Qy	1281	-----	1281
Db	131892	CGGACTCAACACACACGCGCCAGACAGCCGTTGTGGGTGTCTCCACTTTGCTGAGAGGG	131833
Qy	1281	-----	1281
Db	131832	GCCCTTGGAGTTGGCGTTGTAGCCTGGACAGTTGGGGTCTGGGCCCTAGGTTTCAATGCTC	131773
Qy	1281	-----	1281
Db	131772	AGGTCCTCTATGAGAAATACCTTCTCGCCCTGGTGAGACAAGCAGCCCTTGTAGGG	131713
Qy	1281	-----	1281
Db	131712	CTGTAGGGTCTATAGAGGTCTCTTTAGAACAGGAGAGCTGAAAGTGATATCAGGGTGCA	131653
Qy	1281	-----	1281
Db	131652	GGCAGGGGATGACTTCTCTCTGACGGCATACAACCCAGATGAGGTGGGGCTGGGCTT	131593
Qy	1281	-----	1281
Db	131592	GGCGTTTCCCTTACATAGGCTGGGAAAGCTGGGGCAGCCCTTGTAGTGCCTGACAGGG	131533
Qy	1281	-----	1281
Db	131532	AAGGGCTGTGGCAGAGCATGCCACCTCTTCTCTGGCTAGAAAAAGTGCAGTGTCTAC	131473
Qy	1281	-----	1281
Db	131472	CTAAAGTTGAGGAACCCGCCCTGGGAAGCATCCCCCATTAACCCAGGACCCCTCAACA	131413

QY	1281	-----	1281
Db	131412	CTGGCTCTTTCAGGCTGGGGACTGCCAGCTTCCCTATTCTTTAGCCCCACAGCCCCCT	131353
QY	1281	-----	1281
Db	131352	ATCTCTTGAGGTCAACAGTACCCCGTGTGTCCAGGGAACAGGGCAGCTGCCTCAG	131293
QY	1281	-----	1281
Db	131292	GTGCCAGCTGCATGCTCTCTGTGTACAGCCCCAAGAGCTAGTAGCACTGAGTCACAGCT	131233
QY	1281	-----	1281
Db	131232	TGGAGCCTCCAGAGCTTAGATTGTAACCTTCAGCCTGTCTTTGTGGAGATGGGGTGT	131173
QY	1281	-----	1281
Db	131172	CCTATGCGCTTTTCTGGGCCAGTGGGTCTGAGGCACTTCAGAGTGTCAACCCACCGTCA	131113
QY	1281	-----	1281
Db	131112	GTCCAGCCCCAGCTGCCACCTTTCTTTACAAATATTTCTTTGTAATTCATATAGC	131053
QY	1281	-----	1281
Db	131052	CAGCATGTTCCAGTCAACAGTGATTAACAAGCCACTCTGTCCCTGGCAGCAGAGGCAC	130993
QY	1281	-----	1281
Db	130992	ACCTGTAACTGGCCTCACATGGTCCCTTGGACTCACACATATCCACTTGTGCACATGAC	130933
QY	1281	-----	1281
Db	130932	AGCTGCAGGCGCTTGTGGAGGCTTCCCTGCCGACTCAGACTGTGTGCAATCTTTAGTGT	130873
QY	1281	-----	1281
Db	130872	GCACATACCTATGGGTACAGAGCATACCCACCCTTACAAGCCCGGTGGGCATG	130813
QY	1281	-----	1281
Db	130812	CCCTGCTCTTCTGCCAATTGAGTGGTGTGAGGGTCCCATCTGTGGGATGGCAGGCGCT	130753
QY	1281	-----	1281
Db	130752	CCTTCCAGCTCTGCCCTATTGGCGGCGCTAGCACCGTGCCTGTAGTGCAGCCCCCTTGAC	130693
QY	1282	-----GlyGlyMetSerValThrGlnCysSer	1290
Db	130692	CCGTCCCACATCGTGTGCTTTCTCCGGCTCCAGGTGGTATGTCCGTGTACAGTGTCT	130633
QY	1291	LysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgThr	1310
Db	130632	AAGGAGGATGGAAGGAGCAGCTCGGGCCACCCCATGAGCTGCGGCCCTTAAACGCACC	130573
QY	1311	TyrAspMetMetGluGlyArgValGlyValGlnIleSerAlaSerIleGlu	1328
Db	130572	TATGACATGATGGAGGGCGGTGAGGCAGACTGTGCACCTCAGCCAGCATAGAGGGTAAG	130513
QY	1328	-----	1328
Db	130512	TGTCGCGTGTAGCCATTACCAGGACGAGGCGTGCACCAGAACCTCCAGGACCCAG	130453
QY	1328	-----	1328
Db	130452	GCCAGAAATGAAAGAGAGTGGCTCTTTGAAAGCTTGGAGGATCCAGTTAGTTGG	130393
QY	1328	-----	1328
Db	130392	CTCAATATTAACAGATAGAGAAACTGAGGCACAACTACTGAGAAATGGGGCCAGGAA	130333
QY	1328	-----	1328

Db	130332	GTGGGGGTGCTTTTCTTCTGCTGGTCTTCATCATGATAAATGTTAGTGTAGACCCAC	130273
QY	1329	-----GlyLeuMetGlyArg	1333
Db	130272	AGCCCTCCAGACACCATATGTTGACACATGCTCTTCTCCCCAGGACTCATGGCCGC	130213
QY	1334	AlaIleProGluArgHisSerProHisLeuLysGluGlnHisIleArgGly	1353
Db	130212	GCATC---CCTGAGCAGCAGACGCC---CACCTCAAGGAGCAGCATCATCCGAGGC	130159
QY	1354	SerIleThrGlnGlyIle	1359
Db	130158	TCCATCAGCAAGGTAC-CTCCCTTATGTTGTGTGTTCCGCCCCCGGCCCAAGTACCAG	130100
QY	1359	-----	1359
Db	130099	CTAAGGGGTGAAACAGAGGGTGAGCTTCCCCACCCATGGTCTATAGGTGAGTGAATA	130040
QY	1359	-----	1359
Db	130039	CTTGGCCCCGACTCCAAAGCTTTCAGCAGCAGCCTTGCTAGTGTGGAGGGGTCCA	129980
QY	1359	-----	1359
Db	129979	GATTGAATGGAACCTTATGCTCGTGTCTCTGTGTAACTGACAGTGTGTTTACTTAG	129920
QY	1359	-----	1359
Db	129919	CATCTCAGCGCATATGAAGACAGACAGGCTTAGAAAAGTCAGTTCCGATGCCCCAGGGCCC	129860
QY	1359	-----	1359
Db	129859	ACAGGAAGGAGGATTTGTGTGCTAGCCTAGACCGAGCCAGGCAAGCAGTCGACAGCCAGT	129800
QY	1360	-----ProArgSerTyrValGluAlaGlnIleAspTyrLeuArg	1372
Db	129799	GCTGTGTGCTGACAGGCATCCCGAGTCTCTATGTGGAGCGCAGGAGACTACTTACCG	129740
QY	1373	ArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProProProProPro	1392
Db	129739	CGGAGGCCAAGCTCTTGAAGCAGAGGAGGACACACCCACCCACCCACCCCTCGGAC	129680
QY	1393	LeuThrGluAlaTyrLysThrGln-----AlaLeuGlyProLeuLysLeuLysPro	1409
Db	129679	CTGACTGAGACCTTACAAGCCCGGCCCTCTGGACCTCTGGGTCCCTGAAGCTGAAGCCG	129620
QY	1410	AlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIlePro	1429
Db	129619	ACTCAGAGGGTGTGTAGCACTGTGAAGAGCGGGCGCTCTATCCATGAGATCCCG	129560
QY	1430	ArgGluLeuLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly	1449
Db	129559	AGAGGGAGCTGCGCCGACACCTGAGCTACCCCTGCACACCGGCTCTTGAAGAGGGGT	129500
QY	1450	SerIleThr-----	1452
Db	129499	TCCATCAC-CCAGGTACGAGAGTCAGAGAGGGGGGCATGGGCAAGCCTGGGGGTGCT	129441
QY	1452	-----	1452
Db	129440	ATGCTGCCCTCGTTGATTTGGGTATAAACTGAGGGCCAGAGACATCAGCAGCTGGCCC	129381
QY	1452	-----	1452
Db	129380	TGGTCACTGACAGTATTTGTGAACCAACCAAGTCTTATGTCCCTAACTTGTCACTGAC	129321
QY	1452	-----	1452
Db	129320	TCCCCACGTCTCCACCTGCATGTATGTCAAGCCTTAGAGGCACCTGTCTTCACTTGGT	129261
QY	1452	-----	1452

Db	129260	CATAGCTTGATGGGAAGAGAAAGACAAACAGCCAGACAAAGGGCGCTAGGGAGGGCTA	129201
Qy	1452	-----	1452
Db	129200	TCCCTGAGCCAGGACACACTCAGTGATGGCCAGAAATGATTCTAGAGAATTCTTTTGA	129141
Qy	1452	-----	1452
Db	129140	AAATTTAGTGCCACTCTTCTTTTGGGCCAGCACCTTAGAAGGCTAAGTCGGGAAGGTT	129081
Qy	1452	-----	1452
Db	129080	ATAAATTTGAGGCTAACTAGGCTACAAAGCTGTACTACTAGCTTAATGCGCAGTCAACACA	129021
Qy	1452	-----	1452
Db	129020	GTTATAAGTTTAAATAACAACACAGACATTGAGTAGCAATTGAGATCTAGGACCTAGCTG	128961
Qy	1452	-----	1452
Db	128960	TGCTCCGTGTCCTCGGCTTGATCCCGCCCATCTCTCCATGATGTCCAATAATTCCC	128901
Qy	1452	-----	1452
Db	128900	TTCTCTGTGGTGCCCTCTGAGACCTAACAGGGAATATGTAACTCCAGCTGGGACCCGAG	128841
Qy	1452	-----	1452
Db	128840	CCACAGCTGGCGTGGCAGCAAGTGATCTGCTATGGAATTCATAGCATTTGCAAGATGG	128781
Qy	1452	-----	1452
Db	128780	TGCAGTCCCACTGGGTGAGGAAGTTGTGTGACCTGAAGTTAGCCTCACTCTGGAACC	128721
Qy	1452	-----	1452
Db	128720	CCTGGGCGGTGGGTGACCTCGATGTGTAGCCAGACTCAAAGTGAGCAGGCGCTCCAGCC	128661
Qy	1452	-----	1452
Db	128660	TCTGGACATCATGTCTGCTAAAGGTCTCTGTCCATTGGCCCAAGGATTAGAGATGCC	128601
Qy	1452	-----	1452
Db	128600	CATCTGTGCAGACATACAGGAAGAGCGGCATCTGGCCAGCCAAATACAGACCTGCACCC	128541
Qy	1453	-----GlnGlyThrProLeuLysThrGlyAlaSerThrThrGlySerLysLys	1470
Db	128540	TCTCCCAGGGCACCCCACTCAAGTACGACTCTGGGGCACCCCTCCACTGGCACCAGAAA	128481
Qy	1471	HisAspValArgSerLeuLeuGlySerProGlyArgThrPheProValHisProLeu	1490
Db	128480	CACAGCTGCGCTCATCATCGGACGCCCCCGCGGCTTCCCTGCGCTGCACCCCGCTG	128421
Qy	1491	AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer	1510
Db	128420	GACATAATGGCTGACGCGGGCACCTGGAGCGTGCCTGTATGAAGAGAGTCTGAAGAGC	128361
Qy	1511	ArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleVal	1530
Db	128360	CGGTCCAGGACACGACGAGTGGTGGAGGGGGTCCCATCACAGTGGGGGTCCAGTCTGCTG	128301
Qy	1531	ProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPhe	1550
Db	128300	CCTGAACTGGGCAAGCCACGCAAGCCACTGACTTACGAAGACACCGGGGACCCCTTC	128241
Qy	1551	AlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGln	1570
Db	128240	ACCAGTCACTGCCACGCTGGCTCCCTGTGTACCAAGAGGAGGCCACGCCAGCGCTTCAG	128181
Qy	1571	Glu-----	1571
Db	128180	GAAGGTGAGTGGGTGAGACTCCCTCAGGCTGCCAGCTGTGGGGAGGGGACAGCTCACTTC	128121
Qy	1571	-----	1571
Db	128120	CTACAGGATACCTGGGCCACTGCTCTTGGCCAGCAGTCTGTCTCTGCTCCCTGCTCT	128061
Qy	1571	-----	1571
Db	128060	ACTGTCATACAGGGCCCCCTGAAAAAGTGTTTTTTAAATAAATTAAGGCTGAGGAGAAG	128001
Qy	1571	-----	1571
Db	128000	GCTCAGTAGTGCNTAAGCATGAAGACCAGAGTTCAAAATCCCCAGACCCACATAAGCCA	127941
Qy	1571	-----	1571
Db	127940	GGCACACTCATGCAGCTGTGTAATTTCTATGCTTTCACAGTGAGAGGAGGAGGACAG	127881
Qy	1571	-----	1571
Db	127880	GAGAGTCTCCTAAAGGTCTCGGGCCATAAGTGGGAGTTAAATATCCAGACATGTTGGAG	127821
Qy	1571	-----	1571
Db	127820	AGTGAAAACTGACCTTCAAAATTTGTCTTGACTTCTCACTACACACCATGGCACATTTCTC	127761
Qy	1571	-----	1571
Db	127760	CATTTCATGTGCACACACATACACAGAAAAATTTAAAGTAAAAATCGTTTGAATGAAACAG	127701
Qy	1571	-----	1571
Db	127700	TTATCACTCCGTAGGCTTCATCTGTCCAGTAGCACCTGTGTGTCTGAAGCCCTCGCTGTG	127641
Qy	1571	-----	1571
Db	127640	ACGTACCCCGAGTTCTGTTCCTCTCTGATTTCTTTTACCTCCCTGACTTCCACTCT	127581
Qy	1571	-----	1571
Db	127580	CATTCTCGATCCTCTCAGCCTTTCCCTGTGTGGCCCCCTCTTTAGAGGGCTCCACCT	127521
Qy	1571	-----	1571
Db	127520	TTGGTTCTGTCTTAGCCAAACCCCTGTATATACCCCAACACCCCTTAATTCAGATGTCAGGG	127461
Qy	1571	-----	1571
Db	127460	CCACCAGGGAGGCTATGGCCAGTTCCAGCCACATCTCTGGCTAGGCTTTGTTAGGTGTC	127401
Qy	1571	-----	1571
Db	127400	AGATCTGTGTTTGATTTGATTTGATTCATTTGTCAGGGGGGACCTTCCCAAGCATTTGATCATTTCT	127341
Qy	1571	-----	1571
Db	127340	CTCTCAAGCCTGTGTCTTCTTCTTGTGTTTCTTCTCCGACCTGAGCAAAATGGCTCC	127281
Qy	1571	-----	1571
Db	127280	TTGGGGCTCAGGGTTCTTGGCTGCTTTGTTCTTGGGGTGTCTCCAGCCCATGATAAAGTCA	127221
Qy	1571	-----	1571
Db	127220	GACTGTGGAGTGAATCGGGAATGGTGACCGAGGAGTAAATGGATATTTTCCAGGGGAC	127161
Qy	1571	-----	1571
Db	127160	CAACAGCAGAGGGGCAAGAGGCCCTGGGGTTGTGACAGCTCAGCCTGGTGGCCCCCACTG	127101
Qy	1572	-----GlySerLeuSerSerSerLysAlaSerGln	1581
Db	127100	AGGCTGAGGTACTGTGACCTGTCCCCACAGGGCAGCTCTCTATCCCAAGGCGTCCAG	127041

Qy 1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro 1601
Db 127040 GACCGAGGTGACATCTACACCCCGGAGATGCCAGTCCCAACACAGCACTGTGCC 126981
Qy 1602 GluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyVal 1621
Db 126980 GAGCACCACCTCACCCCATCTCCCTATGAGCACTTGTCTCCGGCGGTGACTGTGTG 126921
Qy 1622 AspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIle 1641
Db 126920 GACCTGTACCGTGTGCATCCATCCATTTGGCCCTTTGACCCCACTCCATACCCCGAGGATC 126861
Qy 1642 ProLeuAsp 1644
Db 126860 CCTCTGGAAGACAGTGGGTGTCTGGGCTGTAGATTGGGGCTTAATGATTCTACTTCC 126801
Qy 1644 1644
Db 126800 TTGGGTGGGCTCTGGGGTCTGGAGAGTTGGGGCTTAGTATAGAGATCATCCCT 126741
Qy 1644 1644
Db 126740 GGTTAGCCAGGCGAGGTTCCAGGGTGGGCTTGGGATCCTGTCTCTCAGTGGGCTTCT 126681
Qy 1644 1644
Db 126680 TTCCTGAGGATTCTCTGTACCTTGCAGTAAGTAGGGGCTCAAGTAGTGCATGACT 126621
Qy 1644 1644
Db 126620 TGAGCCCTTTTAAAGGCTCCTTTGTAGGTGTGAGGAGTATATTTTGGCGATTATAT 126561
Qy 1644 1644
Db 126560 CTGTGAGCACTTCCAAAGCATTTGACAGCAGACACCCACATGATGAGGAGCAAGTCT 126501
Qy 1644 1644
Db 126500 AAGCTCTGTGTGTACTGGGAGGATGTGGGTGCACAGGAGGAAGCTGAAGTGGG 126441
Qy 1645 1647
Db 126440 GAGTATGAAGATAAGACATGGGCCACCTGACCTCCCTCCCTCCCGCAGCAGCGCA 126381
Qy 1648 AlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProPro 1667
Db 126380 GCCTACTACCTGCGCGGCACTTGGCGCCCAAGCCCACTTACCCACACCTGTACCCACCT 126321
Qy 1668 TyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleAsn 1687
Db 126320 TACCTCATCCCGGCTACCTTGACCGGGCCCTGGAGAACCCCGCAGACCATCATCAAT 126261
Qy 1688 AspTyrIleThrSerGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAla 1707
Db 126260 GACTACATCACTCGCAGCAGATGACCAACAGCTGCTCCGCGCATGGCCAGCGTGTCT 126201
Qy 1708 AspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly 1727
Db 126200 GACATGCTGAGGGTGTGTACCGCGAGTCTCGTGGCCCTCAATTATGCGCGTGGC 126141
Qy 1728 ProArg 1729
Db 126140 CCAAGAGTGAGTTACCCAGGACTGGGCCCATCAGCAGAGGCTTGCAAAAGACAGTGTCT 126081
Qy 1729 1729
Db 126080 TTTGTGTCACTGTGTAAAGTGGGAAGAGGGGTACAGTTGCCAGGTGAAGTAGTGGT 126021
Qy 1729 1729
Db 126020 GCATGCTGGCTCTTGGGAGAGCAGGTGTGCTTAGATCTTTTACGCTTAGCAGCTGT 125961
Qy 1729 1729

Db 125960 GTAGGTAGCAGGCACCTCTTGTCTACCATGATGTGTCTGAGACTTTGGTATGAGCAT 125901
Qy 1729 1729
Db 125900 CTTTGTACATTCAGAGATACAGAACACAGGCTGGACCGGGTGTGTCATACCTTTTCTCTA 125841
Qy 1729 1729
Db 125840 AGTATCGGGAGGTTCAAAGCCAGCCTCAGCTACATAAAGCTTGGAGTCCAGCCTGGGC 125781
Qy 1729 1729
Db 125780 TATGTGAGTCCATTTTAAATACAAATGTGTGTGTGTTCCTCCCTCCCAAAAAACAA 125721
Qy 1729 1729
Db 125720 AGAAAGGAATGAGTGAAGACCTTTTGGAGCACTGACTTCCAGAGCCTGTGAAGAGCCCA 125661
Qy 1729 1729
Db 125660 GGTCACTTTGTCAAGACTGTTGAGCAGCTCTGTGGGTTTTCAGAGGGTGTGGCGGTGT 125601
Qy 1729 1729
Db 125600 CGGTTCAAAGATGTGTGTGAGAGACACTACCGGTTCAGAGCCTTGAGAAGGAGCCTGG 125541
Qy 1729 1729
Db 125540 GGCAAGCCTGGCCCTTAGGAGGGGTACAGAGGGTCAAGGCTCAGGCTACACCTTGTG 125481
Qy 1730 1733
Db 125480 GGAACAGGAGGTACAGAACCTTCAGCTAATACCCCTCCGACCTTCAGGCACTTATCGAC 125421
Qy 1734 LeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThr 1753
Db 125420 CTGTCCAAAGTGCACACCTGCGGTGTGTGTGTCACCAACGCGCAGCCCTTGCACAC 125361
Qy 1754 AlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer 1773
Db 125360 GCCATGACCGCTTGCCTACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 125301
Qy 1774 SerSerProLeuSerPro 1779
Db 125300 AGCTCACCGCTGTCCCGCAGTGGCGCTGGCAAGTGGGTGGTGGTGGTGGTGGTGGT 125241
Qy 1779 1779
Db 125240 CGATGCCCTGACAGTAAGTTCTGGCCCTTGGCTCTGGCTCACACTCTGGCCCTGGGGTT 125181
Qy 1780 GlyGlyProThrHisLeuThrIleValProThrThrThrSerSerSerGluArgGluAr 1798
Db 125180 TCTAGAGGCCCCCAGCTACCTAGTAAACCACTGCCAATCTTCATCGAGGCGGAGACG 125121
Qy 1798 GAspArgAspArgGluArgAspArgAspArgGluArgGluArgGluArgGluArgGluArg 1818
Db 125120 GGAACGTGACGGGAACGAGAC 125079
Qy 1818 rThrThrValGluHisAlaProIleThrArgProGly 1830
Db 125078 CACTTACAGTGGAGCATGCACCCATCTGGAGACCTGGTAGGACACCCAGACCCCGCCCCA 125019
Qy 1831 1836
Db 125018 CCCCCAGGTCTCATGGGCCACCTGGCCCTGGCCCTGGCTGACCTCGGGCGCTG-TCACAGG 124960
Qy 1836 ySerSerGlySerSerGlyGlyGlyGlyGlySerSerSerSerSerSerSerSerSerHi 1856
Db 124959 TACGAGCAGCAGCAGCGGGCTGGGGC 124908
Qy 1856 sAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSe 1876

Db	120545	TGGCCATTTACAGGAAATGTGCAGAGGCGCGCTGTGGACAGTGGGGTCTCCTCATCATC	120486
Qy	2205	-----ArgSerProGluProAsnLys	2211
Db	120485	GGGTAGAATTCACCCCTCAGCTAGACGCTATCTCCATCAGGTCCTCCAGAACCCAGCAA	120426
Qy	2212	ThrSerValLeuGlyGlyGlyLeuAspGlyIleGluProValSerProGluGlyMet	2231
Db	120425	ACATCGGTCTCTGGCAGCAGCGAGGATGCCATTGAGCTGTGTCCCCACAGAGGGCATG	120366
Qy	2232	ThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGln	2251
Db	120365	ACTGAGCAGACATGCTCGAGCAGCTCGGTACCTGCTGTATCGAGACGGGAAACAG	120306
Qy	2252	ThrGlu-----	2253
Db	120305	GGCGAGCCAGGTACCATCAGCACAATAATCCGCGCTTAGAGATGGTGTGTCAGTTCCATG	120246
Qy	2253	-----	2253
Db	120245	TACACCAAGGCTTCGGCCCTCCCATATGGAGGCCATTAATCCCTCAGACTCAGCGGC	120186
Qy	2253	-----	2253
Db	120185	TGATAGACCTTCAGGGGCTTCAGTTGGAGCACCTGTGTGTGGCATCTGTGGTCCTGGC	120126
Qy	2253	-----	2253
Db	120125	ATCCCTGTGTGGATCTCAGCCTCAGCCATGCCCTGTTTATCATCTGCTGTGTGATCT	120066
Qy	2253	-----	2253
Db	120065	GGGAATGAATAGTTGAACAGACTGGTGGGGTGGACATGGACTTCAGGAGGGTGT	120006
Qy	2253	-----	2253
Db	120005	GACCTGCACCTGTCCAGTTCCTGGGATTTCTGTGTAGATCTAGCAATAAGGCATTGC	119946
Qy	2253	-----	2253
Db	119945	TGGGGCAATTACAGCGTACTCTGGGATGGGAACCACTGGTCTTAGCACCTCACCGTA	119886
Qy	2253	-----	2253
Db	119885	GCTCAAGTGTGAGTCCCTGTTGCTAAAGTCAATAATGCCAGCTGCTGACCTGGTG	119826
Qy	2254	-----ProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAla	2270
Db	119825	TCTCCTGCTCCCTCGAGGATGGGCTCTAAGTCTCCAGGCAACACCCAGCCAGCCAGCC	119766
Qy	2271	PhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIle	2290
Db	119765	TTCTTCAAGTGAAGTGAAGCACTCCGCCATGTGTGAAGTCAAGAGAGGAGATC	119706
Qy	2291	AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyr	2304
Db	119705	AACAAGAAATCTAACACCCCAACCCGGAACAGCGCAGATACAGTAAGGGGAGCCAGGG	119646
Qy	2304	-----	2304
Db	119645	CTAGAGCAGGACCGGAGGAGATAGTACTCTTTAAACCTTACCATGGAGCCCAACG	119586
Qy	2304	-----	2304
Db	119585	TTCTGCCAGTTGGTACTAACTGACACAGCATATAGCATTGTGTAATGCCCTGCACAC	119526
Qy	2304	-----	2304
Db	119525	GTAGTGACACAGGAAAGTGTAGAGACAGCCCTGTGTGTGGTCTGAGGAGCACAGCAC	119466
Qy	2304	-----	2304
Db	119465	ACTACTCCAGTTTCATGCTGCAGACCCGCGTCTTTTACCACCGGCTGATGTGGGTGGACA	119406
Qy	2304	-----	2304
Db	119405	AGGACTCATGCTCTTAGCGCAGGGTCACATGGGCCATATGCTTACAAGAATGCACCGAA	119346
Qy	2304	-----	2304
Db	119345	AGCAACACATGACAGCTTTGCACTGTCCCTCCAGGGTGGCTGGACAGACAGGCAAG	119286
Qy	2304	-----	2304
Db	119285	GTTCTTCTTCTTGGAGTATAGAAGCCCTCACAGGTAACAGAGSCCTAAGGTAGTGACCA	119226
Qy	2304	-----	2304
Db	119225	ACCGCTGAGTAAACCTGGAGCATCAGGTGCTGTTGTGCCAGACATTCACACTCAGAC	119166
Qy	2304	-----	2304
Db	119165	TGTACGGGGACAAGCAGGGGTGCAGATCTCTTAGACCTGTGTGAGCTCAGCCCTTCCCA	119106
Qy	2305	-----AsnIleSerGlnProGlyThrGluIlePheAsnMetProAla	2318
Db	119105	CCTTGTACCTGCTCTAGATATTTGCCAGCCTGGGACGGAAATCTTCAACATGCCCGCC	119046
Qy	2319	IleThrGlyThr-----	2322
Db	119045	ATCACTGGAGCAGGTAACCTCCTGTGTGTCGCGATGGCAAACTCCCTTCCCCAGTGTG	118986
Qy	2322	-----	2322
Db	118985	ACTTGGGACAAGAGGCTTTAGCTCTCTCCCTGTCTCAGTGGACCCAGAAATGACAGAT	118926
Qy	2322	-----	2322
Db	118925	GGGAGCCTCCACTGCTCCTAGGTCCTCCCTACATGAGGAGTGCCCAAGTTTGAAGTCA	118866
Qy	2322	-----	2322
Db	118865	GGTCCCTCAAGTGGTGTAGGTACACCCCGGAAACACTTGGCCAGTTGTGGGGCTTGG	118806
Qy	2322	-----	2322
Db	118805	GGCTGGGGCAGCTCACCAAGTGTGACCTGAGGGATTGGCTTGGGGTACAGAGATTCA	118746
Qy	2322	-----	2322
Db	118745	GGTGGCCGAGGATCTTTTACACTAAAGTCTCTGGCCATTTTCTAAATCAGATGCAGGA	118686
Qy	2322	-----	2322
Db	118685	GTAATGTAGGGAAAGTGTCAAACTCACGTCGAGTTTGTGGCAGAAAGTCTCAGGCTTTC	118626
Qy	2322	-----	2322
Db	118625	AGTATGTGTCTGGGGCTTAGAGCATCCGCTTCCGCTGCTTATTATCACGTTGCCCTGT	118566
Qy	2322	-----	2322
Db	118565	GGAAGAGAGCTTGTCTGGGAAACAAATGGGCTCGGTGACTGTGTCAAGGGAAAGTGTG	118506
Qy	2322	-----	2322
Db	118505	GGGACTTAATCCAGGCCAGAGGAACAGGCAAGGTCCTACTTCTTATTTGTGTGGAGC	118446
Qy	2322	-----	2322
Db	118445	CTTGGGCTTAGGGAATAAGGTGGCCAGATTAAACCTTGGGGACATTTCCAGAGACCA	118386
Qy	2322	-----	2322
Db	118385	GGACCTGTAAAGGGATGGGGTAGGATCTGTGGCAGGACAGCCTTGGGGCTTGTCTGAGC	118326

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, W., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpthy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseg, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajd, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (19-Oct-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 254449)

Rat Genome Sequencing Consortium.

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23265645. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GIBP

Center clone name: CH230-157E10

----- Summary Statistics

Assembly program: Atlas 3.0;
 Consensus quality: 245355 bases at least Q40
 Consensus quality: 247474 bases at least Q30
 Consensus quality: 248954 bases at least Q20
 Estimated insert size: 254924; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 254449: contig of 254449 bp in length.

FEATURES

Source

1. .254449

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-157E10"

misc_feature

3075..3548

/notes="clone_boundary"

/site="EcoRI"

end sequence: BH283564"

complement (251005..251579)

misc_feature

/notes="clone_boundary"

/site="EcoRI"

end sequence: BH283551"

ORIGIN

Alignment Scores:

Pred. No.: 2,59e-61 Length: 254449

Score: 3479.50 Matches: 1345

Percent Similarity: 18.72% Conservative: 71

Best Local Similarity: 17.78% Mismatches: 167

Query Match: 26.33% Indels: 5989

DB: 2 Gaps: 43

US-09-522-753-5 (1-2517) x AC097560 (1-254449)

QY 905 ThrAlaLysSerSerglyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAsp 924

Db 234429 AGCACAGGACTGCGCTGCCACCCCTCATCTGGAGTAGTCCCGTCCGACCATCCCTT 234370

QY 925 GluValAspGluAlaGluGlyAspLysAsnArgLeuSerProArgProSerLeu 944

Db 234369 GCCATAGACCACAGTCACGAGGTGACGAA-----GCCTGACCAGCAGCT 234325

QY 945 LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys 964

Db 234324 CTTGGTCT-----GGACACCCCAATTGGTGACATTCGCTGGGAG 234286

QY 965 GlnLeuLysGlnArgAlaAla-----AlaIleProIleGlnValThr 979

Db 234285 ACCGTGAAGCAAGATAGTCCCTCTCACACATGCTCTGTCTGCTGCTGCTGCTG 234226

QY 980 LysValHisGluProProArg-----GluAspAlaAlaProThrLys 993

Db 234225 TCAGCCACAAGGAGGCCCCATACCAAGCAGTTATATATGGGACATACACACCTTGTCTG 234166

QY 994 ProAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGln 1013

Db 234165 ACATGTCCCAATGGCCCTCATGTCTAGGATACTGTCTAGCTGACCATGAGATGTCTAG 234106

Db	231949	GCTCCACTGCCAGTGAAGCATCACCAAGGGCCTCCCGACTCCCGACTGCAGATGGCC	231890
Qy	1217	-----	1217
Db	231889	CCAGTACCGAGGCTCTATACCCACGTAGTGTCTGGGGCAGGGTGGATGGGCAGGCA	231830
Qy	1217	-----	1217
Db	231829	TCCTAGGAGGTTGGGATGGGATGAGAGGACCTCCGGGGAGACGAGATCTTGACCTGA	231770
Qy	1217	-----	1217
Db	231769	ATGTTGGTCACTGCGGTGTGGGTGAGGCTAGGCTAGGTGATGAAGAAAAACAGGTGTCTATGT	231710
Qy	1217	-----	1217
Db	231709	GTCTCAGCAGTTGGGTGTGGCAGAGACCAACACAGAAAGGATAGCCAGGCCCTTAACC	231650
Qy	1217	-----	1217
Db	231649	AGACTGTCAGGAGCCATAGATCATAAAGTGTCTGGACTGAGAAAGTTACCAAGGTGACCCC	231590
Qy	1217	-----	1217
Db	231589	AGTGACATGTTGCCATGGCCAGAAAGACAGAGGGGACAGAGGTGGCAGCTGTCTTGA	231530
Qy	1217	-----	1217
Db	231529	CAAAATCACCAGTCCGTAGGCATGAGCGAGTGGGTACAGGCTTAGTACCCACGGTGT	231470
Qy	1217	-----	1217
Db	231469	GTGGCCNAGCATACAGCCACGGAACCTTCCAGTCACGGACCACCACACAGAGTCCCTT	231410
Qy	1217	-----	1217
Db	231409	TTGACTCACTGTAGGGAGCAAAATGTCCATTCATGCGCAGGCTTCCCTGAACTCTCTGG	231350
Qy	1217	-----	1217
Db	231349	AAATCTTGGGCTTTCAAAGGCAAGCTCAGAACTGGAGGCGAAATTCAGTCATGTCT	231290
Qy	1217	-----	1217
Db	231289	TCGGCAGTGCAGTAGCCCTCCCTGACTCAGTCTTCCCTGTGACAGAAATGTTCTTCC	231230
Qy	1217	-----	1217
Db	231229	AGGATAGGTACTCCAAGGAGCTCTTAGAGTGCCAGGTTTGCACACAGAGCTTCGCAGGGC	231170
Qy	1217	-----	1217
Db	231169	TCAGGGATGCCGGCTCAGGCTGAGAACTCCCTTCCCTGGTTAGGGCTTGTGGAGAT	231110
Qy	1217	-----	1217
Db	231109	GAGTCAAGCCTCTGGGGCCTATCTGTGTGGATGATGATGCCAGGCATTGATGACACG	231050
Qy	1217	-----	1217
Db	231049	GCTAACAGGTGGAGACACACATCCCACTGTGCTGACCGCCTGGAAACCAAGGCCAC	230990
Qy	1217	-----	1217
Db	230989	AGAGCCCAATGTCCCACTCAGCTCCAGGATTAACACTCAGGCCTTGGAGCGGCTCGGG	230930
Qy	1217	-----	1217
Db	230929	TGGGCCCTGTGGCTCCCTGTAATCTTTGGTTTGTCCAGTGAGTTATACGGAGCAGAAC	230870
Qy	1217	-----	1217
Db	230869	CATGGCCTGGCTGTATTTTCGGTTGTGACAGAGCCCTCTAAGCTTGGAGACTCAAAACATC	230810
Qy	1217	-----	1217
Db	230809	AATACCCCTTTTTCATCTGGAGCCTGCCTGACCTTTCTCACCGTATACCTGGCTAGCAA	230750
Qy	1217	-----	1217
Db	230749	GCAGCTTACAGCCACTTGAGAAGGCATTTCCCGACACACCCAGAAATCACCCCGCATGCC	230690
Qy	1217	-----	1217
Db	230689	CCACGGGGACCAACCCCAACAGAAAGTCCACGGGAGACAAACATCACTGAATCTGTCTCC	230630
Qy	1217	-----	1217
Db	230629	TGAGGTCAATAATGAGAGAGAACTCAGGATTTGGGGCCCTCTGAGAGCCAAGTTAGAT	230570
Qy	1217	-----	1217
Db	230569	CCTCATAGGATGGATACGTGCTGACTAATCTCAGGAAAGCAGGTGACTTGTATTAAAC	230510
Qy	1217	-----	1217
Db	230509	AGCAGTTAGCAACTTGGGGTGATCTTATCTTTATCCCCCATGGGCTGTCTCAGAAAGCTG	230450
Qy	1217	-----	1217
Db	230449	TACTAATGCCACACTCCGGTTTACAGAGGTCACTAAGTGGATACAATCCCAACAAACC	230390
Qy	1217	-----	1217
Db	230389	GCATGCCGTTGGTTTGAGAACTCTCGGGTGGGAGCTTACTCCCTCTAAATCAGTGT	230330
Qy	1217	-----	1217
Db	230329	CTTAGTTGCCCGACCTGCCCCAGGACATCAAGGGGCAATTAAGGGTGTGTTTGGTGT	230270
Qy	1218	-----ProSerAspSerAla-----	1222
Db	230269	TATAGTGGACAGGTTGGAAGGGCGTGAGTCACCTAGTGTATCCACACCCAGGAAGAGTG	230210
Qy	1222	-----	1222
Db	230209	CCTGGCTGCTCTACTACCTCAGTTCCTCTGTGTTGGTGGCTCGGACAGCAGTCCAGC	230150
Qy	1223	-----IleThrTyr-----	1225
Db	230149	TCTTGAAGGCTCAGCAGGAAGGCTGGTGTAGTTCCAGGCTAACCTGAGCTATATTG	230090
Qy	1225	-----	1225
Db	230089	CCAGACACCTTGTCTCTATGGAACAGTATTCGGATAGCGTCTCTCGGAGGTACAAGTG	230030
Qy	1225	-----	1225
Db	230029	AGAGACAGGCTGTTGCGCAGGTGTGACACTGGGAAGCCAGGATGAGTACGGTGGCAA	229970
Qy	1226	-----ArgGlySerIleThrHis-----	1231
Db	229969	TGGTGGGGGGGAAGTTACACATGCTGTGTATACCTTGTGGGTGGGGGAGTAGG	229910
Qy	1232	-----GlyThrProAlaAspValLeuTyrL-----	1240
Db	229909	TGATGTGTCAGGATCACCCACTCTACCTCCCTATAGGGCAGCCAGCAGAGCTCTTACA	229850
Qy	1240	ysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgG	1260
Db	229849	AGGTTACCATCAGCAGGATCGTTGGTGAGACAGCCAAAGTCGCTTGTATCGGCGCACGAG	229790
Qy	1260	luAspSerLeuProLysGlyHisValIleTyrCluGlyLysLysGlyHisValLeuSerT	1280
Db	229789	AGGATACCTTGCCCAAGGGCCATGTCTATGAGGGCAAGAAAGCCATGTCTCTATFCT	229730

QY	1280	yrGlu-----	1281
Db	229729	ATGAAGTGGGTACAGAGGACAGAGACCATGAGGCACTGGCAGTCAGGTGGGTGGCTCC	229670
QY	1281	-----	1281
Db	229669	TGCTGTCCCTTTCTTCCATTCTAAGGATGTCTTTTCTCAAGACAGACATAGCAAGACT	229610
QY	1281	-----	1281
Db	229609	CCTGAGAGCAATCCCTTCTGCTAGGCTCAGCAGGAGACATGTCAGACAGTACCTCC	229550
QY	1281	-----	1281
Db	229549	CTCCAGGTCAGGTGGCTGCTGCAAGGAGTGTGTGGAAACCCCTTGTGTAAGGAGTC	229490
QY	1281	-----	1281
Db	229489	AGGAAAACAGACTGTAAACATGTGTCCCAAGCAGTGTGTGTCTATACACTTGC	229430
QY	1281	-----	1281
Db	229429	TGAGAGGGAAGTCTTTTGGCAGAGCAGAAAGCAGCCAGGGCCTTTGGGTTGGTGTG	229370
QY	1281	-----	1281
Db	229369	TGCTAGCCTCAACAGGTTGAGGACTGGGCCCTAGGCCCTGCTGCCACTTAGACATAGGG	229310
QY	1281	-----	1281
Db	229309	AAACGGGAGATGACATAGTCAATTCAGTGGCTTCTATGAGGTTACCATCTGGCCC	229250
QY	1281	-----	1281
Db	229249	TGCTTAGACAGGCAACCTTGTGTAGGAGAGCCATGGATCTATAAAGTGTCTTT	229190
QY	1281	-----	1281
Db	229189	AGACAAAGGCAGAGCTGAAGGACACCAAGGAGCAGGAGAGTGACTTCTCTCTCTGC	229130
QY	1281	-----	1281
Db	229129	AGGCATACGCCAGAAAGGGGTGAGGGCTGGCCTTGGTGGCTTCCCTACAGAGGTGGG	229070
QY	1281	-----	1281
Db	229069	AAGCTGGGGAATCTTTGAGTGGCTGACAGGAAAGGGCTGTGGTAGAGCCATGACACC	229010
QY	1281	-----	1281
Db	229009	TTCTCTCTGGCTAGAAAAGGAGCCAGGGCTGCCCTAAAAGTTGAGGAACCCATACCTGGGA	228950
QY	1281	-----	1281
Db	228949	TGCTGTCTCCAGTACCAAGACCCACATCAACTGCTGCCATGACTACTGACGGCTG	228890
QY	1281	-----	1281
Db	228889	GGGGCTCCAGCTTCCCTATTCTTGTAGTCTCAGAGCCCTGGTCTCTAGAGTCAAG	228830
QY	1281	-----	1281
Db	228829	TGGCCCCATGGCTTTGGCACCTGACAATAAGTACAGTCTTAAGGACAGGGCTACAGC	228770
QY	1281	-----	1281
Db	228769	TGACCCCGGTGCTGGCTGCATGTCCAGGTGACAGACCCCAAGAGCAGAGAGGGGTGAT	228710
QY	1281	-----	1281
Db	228709	CAGAGTCTTGGCCAGAGCCCTGAGTCAAGGAGTGGAGGCTCTAGAGTCTCAGGTTGTA	228650
QY	1281	-----	1281

Db	228649	ACTTTACAGCTATCTTGGTGGAGACTGGGTATCCTGTGTCTCTTTTGGGGCCAGTGG	228590
QY	1281	-----	1281
Db	228589	GTCTGAAGCAGTTTCAGAGTAGCGCCCAACCGTCAGGCCCCAGCCAGCTGCCAGCCCTTTC	228530
QY	1281	-----	1281
Db	228529	TTGCGCAATCTCCATAGCCAGAGCATCCAAAGTCATTGGGGATAGCAATGCCACTCTGTCC	228470
QY	1281	-----	1281
Db	228469	CGACACGCTGAGGCTTTCATGTCTTGGACCCGGTGTCTACATCCCTTGTGTACATA	228410
QY	1281	-----	1281
Db	228409	TGCAGACCTGCAGGCCCTTGTGGGGCTGTCTGCGGACTCAGACCGTGTGACATCTTC	228350
QY	1281	-----	1281
Db	228349	GTTAGTGTGCACATACACCATGGGCACAGACACATACACCCCTCCTACACAGAGTGG	228290
QY	1281	-----	1281
Db	228289	CAGGCATGCTGGCACTTTCTGCCAAGAGCAGTTCCTCAGTTGAGCTGGTTTGAGGGTCCC	228230
QY	1281	-----	1281
Db	228229	ATCTGGGGGCGAGGCCCTCTTTCAGTCTGTCTGCTTTGCTTAACTCCATGCTGTGGT	228170
QY	1282	-----GlyGlyMetSer	1285
Db	228169	GTCTGCACCTTGTGTCACCCACACCATGTGCTTTTCTGGCCCTCAGGTGTGTGTGG	228110
QY	1286	ValThrGlnCysSerlysgluaspGlyArgSerSerSerGlyProProHisGluThrAla	1305
Db	228109	GTGTCAAGTGTCTCAAGGAGGATGGAGGAGCAGCTCAGGCCACCCCAAGAGTGGCC	228050
QY	1306	AlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAla	1325
Db	228049	GCCCTTAAACGACCTATGACATGATGGAGGCCGTGTGGGAGGACCATTCACTCAGCC	227990
QY	1326	SerIleGluGlyLeu-----	1330
Db	227989	AGCATCGAGGGTAA-GTGTCTCACTGTAGTGTATTACCGGGGTCAAGCACCATCCAGGA	227931
QY	1330	-----	1330
Db	227930	GACCCAGGTTACCCAGGTAAGAAATGGAAAGGGAGGGTGGCCCTTGACATGCTGGAGGG	227871
QY	1330	-----	1330
Db	227870	AGCCAGTCAGTAGTGGCTCAATATTAAACAGTAGAGAACTAGGACACAAACCGGTGCA	227811
QY	1330	-----	1330
Db	227810	AAATGACGCCAGGGAACCGGGCACTTGTCTGCTGGTGTAAAGTCACAAACAGAGACTG	227751
QY	1330	-----	1330
Db	227750	TTGGTGTAGAGCCAGCCCTCTCTGGACACCATACGGTGCACACTATCCCTGTCTCCCA	227691
QY	1331	-----MetClyArgAlaIleProProGluArgHisSerProHisIleuLyGluGln	1348
Db	227690	GGACTATGGGCGTGGCCATC---CTGAGCAGACACAGCCCC---CACCTCAAGGAGCAA	227637
QY	1349	HisIleArgGlySerIleThrGln-----	1357
Db	227636	CATCATCCGAGGCTCCATCAGCAAGGTATATCTCTTACGGTGTGTAAATTCACCT	227577
QY	1357	-----	1357

Db	227576	CCACGCCAAACCAAGCTTAAGGGGTGGAAACCGAGGGTGAACCTTCCGTCCTGTTCTTAT	227517
QY	1357	-----	1357
Db	227516	AGGTGGAGTCACTACTCTGTAATTGGGGCTCATTCMAAGCCCATGCTGCATGCTTGCT	227457
QY	1357	-----	1357
Db	227456	AGTGATTGAGAGGGGTCAGATTGAATGTGTTTGTAGTCTATGCTTGTGTAAACTG	227397
QY	1357	-----	1357
Db	227396	AGCAATGTGTTTCACTTAGCGTCTCACGTCATATGAGAAACAGGCTTAGAAAAAGTCA	227337
QY	1357	-----	1357
Db	227336	GTTTGCATACCCAAAGTCCACAGGAAGATTGTGTGACCCAGGCTGAGCCAGGCAAG	227277
QY	1358	-----	1367
Db	227276	GCAGTGGAGAGCCAATGATGTTTGGCCGCGCAGGCATCCCGAGGTCTATGTGGAGGCCCA	227217
QY	1367	nGluAspTyrLeuArgGluAlaLysLeuLeuLysArgGluGlyThrProProPro	1387
Db	227216	GGAGGACTACTTACGTAGGAGGCCAAGCTCTTGAAGCGAAGAGGCAAGCCACCGCCCC	227157
QY	1387	oProProSerArgAspLeuThrGluAlaTyrLysThrGln-----AlaLeuGlyPr	1404
Db	227156	ACCGCCACCTCGGAGCTGACTGAGACCTACAAGCCCGGCCCTCGACCCCTTGGGTCC	227097
QY	1404	oLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgse	1424
Db	227096	CCTGAAGCTGAAGCCAACTCATGAGGGTGGTAGGACTGTGAAGAGCGCGCGCGCTC	227037
QY	1424	rIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProAr	1444
Db	227036	TATCCATGAGATCCACGAGAGAGGTGCGCCGACGCGCCGAGCTGCCCCCTGGCACCAAG	226977
QY	1444	gProLeuLysGluGlySerilethr-----	1452
Db	226976	GCCTCTGAAGAGGGTTTCCATCAC-CCAGGTACGAGAGCACAAAGAGGGGCGGGGCT	226918
QY	1452	-----	1452
Db	226917	TGGGCAAGCCTGAGGGAGTTCTAATGCTGCCCTCATTGCCATGGGTTATATAACTAAGGCC	226858
QY	1452	-----	1452
Db	226857	CAGAGAGGTCAGCGGCTTGCTCTGGTCACTGACAGGGCTTGTGAAACCAAAGTCTTTTGG	226798
QY	1452	-----	1452
Db	226797	TCATTGGCTGCTCCCOACANTGCTACCTGCATGTTGTGTGTCAGACTAGAGCCACTGT	226738
QY	1452	-----	1452
Db	226737	CCCTTACCTTGGTCATGGCTTGATGGAAGAAAGAACAAACAGCCAGCAAGGCACCTTA	226678
QY	1452	-----	1452
Db	226677	GGGAAGGCTGTCTGTACCTGAGCCAGGGACACACAGTGTATGGACCAAGATGATTCATAGG	226618
QY	1452	-----	1452
Db	226617	GTATCTTTTGAACACTTAGGTGCTGCTCCTCATTTTGGGCCAGCACTTGGGAGGGGAG	226558
QY	1452	-----	1452
Db	226557	TCAGGAAGATTGTAATTTGAGGCTAGCCTAGGCTACAAAGCTGTACTCTGGCTTACTGC	226498
QY	1452	-----	1452
Db	226497	CGATCACCAAGTTTTTACAAATTAATGACAAACAAACATTGAGTGAAATTGAGATCTGG	226438
QY	1452	-----	1452
QY	1452	-----	1452
Db	226437	GACCTAGCTGTGCTCACTGTCCCGCAGCTGGATCCCTGTGGGATCCTCTCCAAGATGTCC	226378
QY	1452	-----	1452
Db	226377	AATAATTCCCGTCTCCGTGGTGTCTCTGAGAGCAACAGGAAATGTAACTCCAGCCT	226318
QY	1452	-----	1452
Db	226317	GGGACCCGAGCCAGAGCTGGGGTGAACAGATGATCTGATATAGAGTCCATAGTGCT	226258
QY	1452	-----	1452
Db	226257	GCAGGATGTCGGGGGCCACACTGGGTGAGGAAGTCAATTGACCTGTAGTACCAAGC	226198
QY	1452	-----	1452
Db	226197	TCTCGTACCCCTCCACAGTGGGGGTCTCTGGACGTATAGCCAGGCTGAAGCAGCAGGG	226138
QY	1452	-----	1452
Db	226137	TATCTAGCCTCTGACGTCTGTCTTCTCAAGGGGTTAGAGATGCCCCATCTCTGCAGAC	226078
QY	1453	-----	1455
Db	226077	ATACAGGAGAAACAGCATCTGGCCAGCCACATACAGACCTGGACCCCTCTCCCAGGGGAC	226018
QY	1455	rProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgse	1475
Db	226017	CCCACTCAAGTAGCACTCCGGGGTGCCTCCAGTGGCTCAAGAAACACGACGTACGCTC	225958
QY	1475	rLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaas	1495
Db	225957	CATCATCGCAGCCCGCGCGGCTTTCTCTGCCCTGCACCCCACTGGACATAATGGCTGA	225898
QY	1495	pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAl	1515
Db	225897	TGCCCCGGGCACTAGACGGCGCTGTCTATGAAGAGCTCTGAAGAGCCGCTCAGGACCA	225838
QY	1515	aSerSerSerGlySerileAlaArgGlyAlaProValIleValProGluLeuGlyL	1535
Db	225837	CAGTGTGTCAGGGGGCTCCATCACCGGGGGCCCGAGTTGTCTGCTGAGTAGGCA	225778
QY	1535	sProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr	1555
Db	225777	GCCCCGGCAAGCCCACTGGCTTACGAGGACCAACCGGGGACCCCTTCCAGCCACCTGCC	225718
QY	1555	oArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSe	1575
Db	225717	CCGAGGCTCACCTGTGACTACCAAGGAGCCACGCCACCGCTTCAGGAAGGTGAGGC-TC	225659
QY	1575	r-----	1575
Db	225658	CCCTGGGCTGCCAGCTGTGGGGAGGGGACAGCTCATTTTCCACCGCCAGTCTGTCTGCTG	225599
QY	1575	-----	1575
Db	225598	CCCTGCTCCACATGCATACAGAGCCCCCGTCTGTTCCCTTACCAATATTTTTAA	225539
QY	1575	-----	1575
Db	225538	TAATTCAAATCAGGGCTGAGGAGGGGCTCAGTAGTAAGTCTTCTGTACCAGCATGA	225479
QY	1575	-----	1575
Db	225478	GACTCAGAGTTTCAATCCCGGTACCCACATTAAGCCAGGCACACTCATACACATCTGTAA	225419
QY	1575	-----	1575
Db	225418	TTCTCATGCTTCAAGTGAACAAAGGCTAAGACAGAAATCTCTCTAAAGTTTAGGCCA	225359

Qy	1575	-----	1575
Db	225358	TGAGTCGGGTCTCAAAACATGGCGAAAGTGGATGACACCTCTCCCCCCCCCTCTATT	225299
Qy	1575	-----	1575
Db	225298	TTATACACGCATACGCACACTCATATCAGCACACTCGCGCTACATTCTCTGTGCAC	225239
Qy	1575	-----	1575
Db	225238	ACATGTACACAAAAACAATAGTTTAAATGAAGCAGTTCACACTCTGTAGGCTTCACCT	225179
Qy	1575	-----	1575
Db	225178	CTCCAGTAGGCTGCTGGTCTGAAGCACCTGCCTGTGACGTGACCGCGCATCCTACTCT	225119
Qy	1575	-----	1575
Db	225118	GAATCCGCTTCTGATTCTCTGCATTTCTGACCTCCGCTCTTTCTCGATCCTCCACAT	225059
Qy	1575	-----	1575
Db	225058	GTGACGCTTTCCGTGTGCTGGCCCTCTATGCCGAAGCTCCACCTTTGGTTTGTCTCT	224999
Qy	1575	-----	1575
Db	224998	AGCCAACCACTGTATATACCCACAGTCACAGATGTGAGGCTGAGAGGCTATAGGCACC	224939
Qy	1575	-----	1575
Db	224938	TCCAACTAGTTCACGCACATCCGGCGGGAGTTGTTTAGGGTGTGACAGCAGTGTCTT	224879
Qy	1575	-----	1575
Db	224878	AATTCATTGTATGAGGAGCCCTTCCAGAGCACTGAGATCCTCAACATCTCTCTCTCT	224819
Qy	1575	-----	1575
Db	224818	CAGGCCGTGACATTCTCTACCTGGGTTCATCTCTATCTGGAACAATGGCTCCTGGAG	224759
Qy	1575	-----	1575
Db	224758	CTCAGGTTCTTGCTGCTTTCTCTCAGGGTGCTCCTCTCTCTCATGGGAATGCC	224699
Qy	1575	-----	1575
Db	224698	AGCCCATGATTTAAGTCAGACTGGAATGAATAGGGCAATGGTGACTGGGAGTAATGAA	224639
Qy	1575	-----	1575
Db	224638	TGCTTTCAAGAGGCCAGGACGAGAGAGACAAGAGCTCTAGGGCTTGTGACAGCTCAG	224579
Qy	1576	-----	Se 1576
Db	224578	CCTGGTGGCCCGTTGAGGCTGAGGCCACTTTGACCTGTCCCCACAGGCACTCTCTCTC	224519
Qy	1576	rSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPr	1596
Db	224518	CAGCAAGCGTCCAGAGACCGGAAGCTGACTCGACACCCCGGGAGATTGCCAAGTCCCC	224459
Qy	1596	oHisSerThrValProGluHisProHisProHisProHisProHisProHisLeuLeuAr	1616
Db	224458	ACATAGCACTGTGCCGAGCACCAACCCCACTCTCCCCCTACGAGCACTGTCTCCG	224399
Qy	1616	gGlyValSerGlyValAspLeuThrArgSerHisIleProLeuAlaPheAspProThrSe	1636
Db	224398	GGGTGTGACTGGTGTGACCTGTACCGTGGCCACATCCATTCGCTTTGACCCCACTC	224339
Qy	1636	rIleProArgGlyIleProLeuAsp-----	1644
Db	224338	CATACCCGAGGATCTCTCTGGAAAGCAGGTGGGTGCTGGGCTGTACATGTGGGGCTT	224279
Qy	1644	-----	1644

Db	224278	ATTGATTGTCTCTCTTGGGTGGACACTTGGGGTATGGCTTAGTGCTGCCATCACCAGCT	224219
Qy	1644	-----	1644
Db	224218	GCTTACCCAGGGCAGGTTCTAGGATTTGGGCTTTTGTGCGTGTCTTCTACCGTAGGACT	224159
Qy	1644	-----	1644
Db	224158	CATCTGTCCCTTGCAGTGCAGTAAATGAGGGCTGCAGTGCCAGTGCCATGATGTGGAACC	224099
Qy	1644	-----	1644
Db	224098	TTTTTAAAGGCTCTCTTTGTGAAGGTGTAGCACCTGTATTTTCGATGCATTATATTGGGG	224039
Qy	1644	-----	1644
Db	224038	GCATCTCCGAAACATTAAAGCAGGCCCCCAACATGATGAGGGAGCGAGTCTTTAAGCTC	223979
Qy	1644	-----	1644
Db	223978	TGTGTGTGCATTGGGGAGGCTGTGGGTGTACAAGGAGAGGAAGCTGTGTGGCAGAAGT	223919
Qy	1645	-----	Al 1645
Db	223918	GAGGAACGTGAAGATAAGTCAATGGGCCACCTGACCTCCATCCCTCCCTCCCTCCAGC	223859
Qy	1645	aAlaAlaAlaTyrrTyrrleuProArgHisLeuAlaProAsnProThrTyrrProHisLeuTy	1665
Db	223858	AGCGCAGGCTCTACTACCTGCCCGGACCTTGGCCCGCCACCTTACCACACCTGTGA	223799
Qy	1665	rProProTyrrLeuIleArgGlyTyrrProAspThrAlaAlaLeuGluAsnArgGlnThrIle	1685
Db	223798	CCGCGCTTACTCATCCGCGGCTACCTTGACACCGCGGCCCTGGAGAACCGCCAGACCAT	223739
Qy	1685	eIleAsnAspTyrrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGl	1705
Db	223738	CATCAATGACTACATCACTCGCAGCAGATGCACCAACGACGCTCTGCCATGGGCCA	223679
Qy	1705	nArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrrAl	1725
Db	223678	GCOCGCGATATCTGAGGGTCTGTACCCGAGAGAGTCTCGCTGGCCCTCAATTATGC	223619
Qy	1725	aAlaGlyProArg-----	1729
Db	223618	CGTGGCCCGCTGGTGAGTTGCTGGGGCTGTGCCCATCAGCAGAGACTTGTCTCTGGC	223559
Qy	1729	-----	1729
Db	223558	TGTGGTGTGCAAAAAGAGTGCACACGTGTGCACCTGCGAGCTGGGCGGAGGACAGTT	223499
Qy	1729	-----	1729
Db	223498	GGCTGGAGGACTAGTGGTGTGCATGCTGGCTCTTGGGGAAAGGGCAGAGCGGGCCACTT	223439
Qy	1729	-----	1729
Db	223438	GTATCTGTTACATCTAGAACTGTGCTTAGCAGACACTCTTGGCCACCATGCACCGTGGT	223379
Qy	1729	-----	1729
Db	223378	CTTTGGTGTGTTCTGMAACCTTGGTATGGGCACTCTTCTGTATCCAAAGAGATCAAAA	223319
Qy	1729	-----	1729
Db	223318	TACAGCTGGATGTGGCAATTGACCGCTATATCTTAAGTACGAGGAAGGTGCAAGGCCAG	223259
Qy	1729	-----	1729
Db	223258	CCTCAGCTACATAAAATTTTGGAGCCAGCCTGGGCTATATAGTCTGCCTCTATACAATG	223199
Qy	1729	-----	1729

Db 218862 CCACAGCCAGACAGATGCTTAACGTTAACTGGCCACCAGCTTAGTCTTTGGGGT 218803
QY 2140 ----- 2140
Db 218802 CATATAAGGTCTAACAGGGACACTTACCCAGGAGACCAAGAGCCAGATCAGGGTTCTG 218743
QY 2141 -----GluValIleThrGlnAsp 2146
Db 218742 CCTACTACTCTGAAGTCTCTGAACTGGCTCTCCCTGAGGAGGTCAATACACAGGAC 218693
QY 2147 TyrThrArgHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrsPhe 2166
Db 218682 TACACAGCCACACACACAGCAGCTCAGTGGCCCCCTTCCCGCCCTCTCTACTCCTTC 218623
QY 2167 ProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrsPro 2186
Db 218622 CCTGAGCCAGCTGCCCTGCTCTGGATCTCCCGCGTCCACCCAGTACCTCTACCTCCCA 218563
QY 2187 ProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyLys----- 2204
Db 218562 CCCCCGACCATGGCACCCCGCCAGCCCGGGATCCCCCACACAGTGAAGGGGGCAAAAGGTGA 218503
QY 2204 ----- 2204
Db 218502 GGAAGTACCATGCTGTCATCAGGTGCTGGGAACAGTGGGACAGGGGTGAGCTTTGCAGT 218443
QY 2204 ----- 2204
Db 218442 GTGGGCTGATGGAAACTGAGGTGACGGGATCAAAAGCCTCTACGCTCCAGAACAG 218383
QY 2204 ----- 2204
Db 218382 ACACAAAGCATCTAATCAAGAAAGTGGCTCACCCAAAGGTCCACCGTGCCAAATTGAATGT 218323
QY 2204 ----- 2204
Db 218322 CCTATATACCAACTCAGCTAGGGGTGAGATGTCCTCACTTCCATTTAGCAGATAGGAA 218263
QY 2204 ----- 2204
Db 218262 ACTGAGGCATAGAGGTGAGTTACTATCCCGTCCCATGACACAAACACCTTGACAG 218203
QY 2204 ----- 2204
Db 218202 TGTAGACAGGTATCTCTGGGTAAACAGCAGGTCTTAGCATGGGGCTCATCAGGGCTTA 218143
QY 2204 ----- 2204
Db 218142 CAACATGGCTCCAAGCACCCAGCAGCACTGTGTGTGGTGGCATACCTCTGGACAGGCTGA 218083
QY 2204 ----- 2204
Db 218082 GGCAAGACGACTGTGAACCTGAGGGCAGCTGGGCTGTGTACCAAGACCCCTGACTCAAAA 218023
QY 2204 ----- 2204
Db 218022 ATAAACCAAGCCCATGAGTTGGGTGAGTTCGAAAGGTCTCAGAAAGCCCTGGGGTC 217963
QY 2204 ----- 2204
Db 217962 TCCATACCTCAGATGCAGAGTCTGGCCATTTACAGGGAAACACGTGAGCCCCACTGTGGG 217903
QY 2204 ----- 2204
Db 217902 CAGTAGGGGTGATCTTTCAATGGAGTAGAACCCCACTCTCAGCTGGGATGCTGTCTCAATT 217843
QY 2205 ArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluPro 2224
Db 217842 AGGTCTCCAGAACCCAGCAAAACA---GTCTGAGTAGCAGCGGATGCTATTGAACCT 217786
QY 2225 ValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrsProLeu 2244
Db 217785 GTGTCCCCACCGGAGGGCATGACTGAGCCAGGACATGACGGGACACTGGGTATCCCCACTG 217726

QY 2245 LeuTyrsArgAspGlyGluGlnThrGlu-ProSer----- 2255
Db 217725 CTGTACCCGGACGGGAGCAGCGGGAGCCAGGTAGCTATGGCACAATAATCCCGCTTAGA 217666
QY 2255 ----- 2255
Db 217665 GGTGGTGGTGTGGTTTCCCATACACCCGTGGCCTCAGCCGTGAGGTGCGTCA 217606
QY 2255 ----- 2255
Db 217605 ATCTCTTACAGCTCAGCGGGCTGATGACTTTTCAGGGGGCTTTCAGTGTGGCACCCTGTG 217546
QY 2255 ----- 2255
Db 217545 CCTGGCATCTCAGCCGAGTCATGTCCTATTATCATTGCTGCTGATTGAGAATGAACCT 217486
QY 2255 ----- 2255
Db 217485 AAAACGGACACAGAGTGGGTAGGGTGGGCATGGAACCTCAGGAGGGGTATGACCTATGGCG 217426
QY 2255 ----- 2255
Db 217425 GCCAGTCCCTGGGAATTCTTGTGTCAGTATAGCTCTAAGAAGTTCTCAGGGGGAATTC 217366
QY 2255 ----- 2255
Db 217365 ACAGTGTGCTGGGAGGCTCACAAACCGCCCTAAGCCAGGGCTAAGTACAGATCCCAGCA 217306
QY 2255 ----- 2255
Db 217305 GGCTGGGATGGAAATGTTGTCCTTAGCACCTACTGTGGCTCAAGTGGTGAAGCTAG 217246
QY 2256 -----Ar 2256
Db 217245 CCCCCGTGTTGTAACACAGTGCACAGCTGCCAGCTGTGACCTGGGCTCTCATGCTCCCTGCGAG 217186
QY 2256 gMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuTh 2276
Db 217185 GATGGTCTTAAGTCTCCAGGCAACACACAGCCAGCCGCGCAGCTTCTTTCAGTAAGCTGAC 217126
QY 2276 rGluSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLysLysLeuAsnTh 2296
Db 217125 CGAGACAACTCCGCCATGTTGAAGTCGAAGAAGCAGGAGATCAACAGAAACTCAACAC 217066
QY 2296 rHisAsnArgAsnGluProGluTyrc----- 2304
Db 217065 CCATAACCCGAATGAGCCTGAATACAGTAAGAGAAAGCCAGCCAGCAGGCACTGGG 217006
QY 2304 ----- 2304
Db 217005 AGACAGAGTGTCTTTAAGCCTTCTGTCGAAGGGCTGATAGTCTGCCAGTGGATACCGA 216946
QY 2304 ----- 2304
Db 216945 CGGCATATAGCACCTCAGTAATGCCCTACGCACAGTAGGTGCTCAGCGAAACCTATAGAG 216886
QY 2304 ----- 2304
Db 216885 ACAGCCCTGTGTGGGCGGAGAGCACAGTGCATCTACTCCAGGTTTCATGCTGCAGGCAT 216826
QY 2304 ----- 2304
Db 216825 GGCGCCTTTTCCACAGCTGATGTGGTGGACAAAGGACCTTGTCTTAGCTCAGGGTCG 216766
QY 2304 ----- 2304
Db 216765 CAGGACCATATGCTTATGAGAATGCATGGGAAACAACACTTGAAGGCTTCGCATGTCT 216706
QY 2304 ----- 2304
Db 216705 CCCTCCCAACAGCAGGCCCTCCCTCAGGTGGCTTGGACAAATCCTTTCTGGAGTTAGACT 216646

Db	214486	ACCCTGGCATTTCAGAAAGTCCAGAGCACAAATTCATGAGAGGAGAGAAAGAAACCG	214427
QY	2322	-----	2322
Db	214426	GAGAGCTTTGAGAGGCTCAGTGAGGCTGCAACATCTCTTGAGGCGAGTGTGTGAT	214367
QY	2322	-----	2322
Db	214366	TTGACTACTCTCTCTGTCTGTCATGGGGTTGGGCCAGCCAGGCTGGAGTGGCCCTTCACCT	214307
QY	2322	-----	2322
Db	214306	GTTAAGCGCTACCTCTCTGTCGACGACATAACTGCATGGCTGTGACCCCCCATCCCT	214247
QY	2322	-----	2322
Db	214246	TGTACCCAGCTCTACCTTTCTCTGCTCTCTGTGGGAGAAATGGGGTGAGGGGGCGAG	214187
QY	2322	-----	2322
Db	214186	CCTAGCTAAGGGGGAATCAGGAGGCTGAGTCCCTAACCCCTGAGAAAGGGGAACNA	214127
QY	2322	-----	2322
Db	214126	CAGGGGGTCTCTCTGGTCACTTTCTGTCGCCACATCCCTTCTAGAACCTACTCACAG	214067
QY	2322	-----	2322
Db	214066	AAGACCCGGGTGAGCTTCTATCCCGTCCCTAGCACCTTAGCTGTGGCTTAACCTCATAG	214007
QY	2322	-----	2322
Db	214006	CACAGCTGCCCTGCCTCTCTTAGACATCCACTTATAGTCAAAATGGAATAATAGAC	213947
QY	2322	-----	2322
Db	213946	CCAGCTGCGATTGGGGTGTAGAGTAGGAACAGGGCTGGGAGCCCTCCGGCTGCCACAG	213887
QY	2322	-----	2322
Db	213886	CCCTCCCTGCAACAGCCATGAGCTGTGGGGCAGTTAGGACCCCATCCGACCTATAC	213827
QY	2322	-----	2322
Db	213826	ACAGAGCGGGCTGGAGGACGACATGCTCGCAGATGTGTCTCCTCTTTC	213767
QY	2322	-----	2322
Db	213766	TCTCTCTCCCTCCTCATTTCTATCTCTCGAAACCTCTGGCTCCCGGCCCGCCGCC	213707
QY	2323	-----	2334
Db	213706	TGTGTGTGTCTGTCTCTCTCCAGGCTTATGACCTGTAGAAGCCAGCGGTGCAAGA	213647
QY	2334	uHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTy	2354
Db	213646	ACAGCCAGACCAACATGGGGCTAGAGGCCATTATTAGAAGGCACTCATGGTAATA	213587
QY	2354	rAspGlnTrpGluSerProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSe	2374
Db	213586	TGATCAGTGGGAAGAGCCCGCGCTCGCGCCCAATGCTTTTAAACCTCTGAATGCCAG	213527
QY	2374	rAlaSerLeuPro--AlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisTh	2393
Db	213526	TGCCAGTGTCCCGCTATGCCATACCACTGTGTGACGAGGAGTACACCC	213467
QY	2393	rLeuThrSerPro-----	2397
Db	213466	ACTCACCTCGCACAGGTCTGCAGGCTGCCCTGCCACCGCCCGCGCTGTGTCCCT	213407
QY	2397	-----	2397
Db	213406	CCTTACCCTGGTGTGTTAATCCTTTTCACTCCCGCCCTCAAGCTTTGGAGAGACTGTGACC	213347
QY	2397	-----	2397
Db	213346	TTATTTTGTGGTGTTTGACATCTAGTTTGTCTAATCAAAGCTGGTGTGACACCCCATG	213287
QY	2397	-----	2397
Db	213286	TTGCAAGGGGCTCCCTCCGCCCCCAGCACTGTCTTTTCCCGTGTCTCACACCACTCC	213227
QY	2397	-----	2397
Db	213226	CCACATCCACCCCTTCCTCCCTGCTGTGTGGACAGGATCAAGCTGACAGGCCACCC	213167
QY	2397	-----	2397
Db	213166	ATGTGCGACCGGGCTACTGTGTGGGAAGAGTCTAGTGTGGGATGACCCCTTGAAGGCCA	213107
QY	2397	-----	2397
Db	213106	GGCAGCACGAGGCCAACTCCCGTTCTTCCAGTGCTGTGTCAGCACTCTCTGTTCCAAGC	213047
QY	2397	-----	2397
Db	213046	AAACAAGCAAGAGAGTGTCTGTCTCCCTTCTAGGGTCAGCCAGCTCAAGGTCT	212987
QY	2397	-----	2397
Db	212986	GGCCCTATCAGCAGAGGCTTTCAGTCTAGCACAAATTGAATGCTGTCTTTGGGCAATGTCTG	212927
QY	2397	-----	2397
Db	212926	TCCTCCAGCAAGATCTGACACTCTCTCAGTGAGGGGTACAGCTGTGGGGCAGCTAAG	212867
QY	2397	-----	2397
Db	212866	GGAAATAGTCAACAACCTCTCTCATAGCAGATGTTTTTCTTTTTCACGCTAGGGCT	212807
QY	2397	-----	2397
Db	212806	GGCTCTGGCCATCTGGTAGACCTCTTAGCCTCTCTGGCTCAGGGTTTCTCACACTAGTTT	212747
QY	2397	-----	2397
Db	212746	CTGTTTGGGGCCTTAGGAGGAGGTGGCTCCCGCAGGAGAAATCGGCCCACTGCCCTC	212687
QY	2397	-----	2397
Db	212686	CTCCGTGTGTGTGGGTGACCCACAGTCAACCATCTAACTCAGCCCTTTCCCACTTT	212627
QY	2398	-----	2414
Db	212626	GTCCCTCTCAGGTGTGTGGGAAGCCAAAGTTTCTGGCAGACCCAGCAGCGGAAAGC	212567
QY	2414	alySerProAlaProGlyLeuAlaSerGlyAspArgProSerValSerValHi	2434
Db	212566	CAAGTCGCGAGCGCCAGGCTGGCACCCGAGAGACGACCTTCTGTCTCTCTCAGTACA	212507
QY	2434	sSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgPr	2454
Db	212506	CTCGAGGGCGACTGCAATCGCCGAACCACTCCCAACCCGTGTGTGGGAGGACCGGCC	212447
QY	2454	oSerSerAla-----	2457
Db	212446	CTCATCTGAGGTGGGTATCAGTAGGGGACAGATGTAGACATGGGGTTCCCTGGGCTA	212387
QY	2457	-----	2457
Db	212386	CCAGGGAGGCGGTCCCTGGTCCCACTAGAAAGAGGGGTGTAGGAAGCCACACATT	212327
QY	2457	-----	2457
Db	212326	GAGAGGTCCTGTAGATCTCTCAGGCCCTTCAGTCAAGGGCTGTGTACACAGGGGGACAG	212267

Qy 2457 ----- 2457

Db 212266 TGCCTGTCACGACGTAGGATTTTACAGAGAAAGCAGAGGACGTAGAGGAAAGTTAC 212207

Qy 2457 ----- 2457

Db 212206 CACAGCCGAGTCAAGGTGGTGGCTGTCTAGCAGAGGCAAGGAGGAGCCGCTT 212147

Qy 2457 ----- 2457

Db 212146 TTGAGAACTCAATGGTGGCTTAGAGGCCAGGCTCGAGCTTCTGTGTGTAGA 212087

Qy 2458 -----G 2458

Db 212086 ATGGGTCTCAGGTGGCTACTTTCAGAGGCCCTCTCCACCCTCTCCCATCTTTCACAG 212027

Qy 2458 lySerThrProPheProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaT 2478

Db 212026 GGTCCAGCCCAATCCCTTACACCCCTTGTATATAGAGCTACAGCGAGTGTATGGCCT 211967

Qy 2478 erProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHisAlaT 2498

Db 211966 CCCACCCCACTGGCTTCCCGCAGCAGCGGGCCCTAGCTGTCTCCACCATGCTT 211907

Qy 2498 rpAspGluGluProLysProLeuLeuCySerGlnTyrGluThrLeuSerAspSerGlu 2517

Db 211906 GGGATGAGGACCCCAAGCCACTGTGTGTTCACAGTACGAGACACTCTCGCAGACGCGAG 211848

RESULT 16

AC136560/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-128N8, WORKING DRAFT SEQUENCE.

ACCESSION AC136560

VERSION AC136560.2 GI:25188528

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE Rattus norvegicus

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 218129)

Muzny,D.,Maré, Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Blawolo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B., Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Cartoll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulsegheh,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,D., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwakoelameh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,

FEATURES

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPVF
Center clone name: CH230-128N8
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208938 bases at least Q40
Consensus quality: 210339 bases at least Q30
Consensus quality: 211081 bases at least Q20
Estimated insert size: 211417; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 218129: contig of 218129 bp in length.
* Location/Qualifiers

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 218129)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (06-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 218129)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 23, 2002 this sequence version replaced gi:24635574.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

source	1. .218129 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-128N8"	Db	80462	GCCTTCCTACACACCCCTGGTGGGACACCCCGATTAGTCTCTGCACTCAG	80403
misc_feature	1. .1421 /note="wgs end extension clone_end:Sp6"	Qy	1080	-----	1080
misc_feature	2131. .3183 /note="wgs end extension clone_end:Sp6"	Db	80402	TGGCAGTCGGCTCTTGGCACTGGCGCAGGGCTGCATGGTCTGGTTGGGTGAGATGAGA	80343
misc_feature	3426. .4108 /note="clone boundary clone_end:Sp6 end sequence:BH284637"	Qy	1080	-----	1080
misc_feature	210619. .211211 /note="clone boundary clone_end:T7 site:EcoRI end sequence:BH284636"	Db	80342	CCTTCCCTGGAGGTTTGGCGAAGCTGGGCACCTCCTGAGAACATCTACACACACCTGTGA	80283
misc_feature	217131. .218129 /note="wgs end extension clone_end:T7"	Qy	1080	-----	1080
ORIGIN		Db	80282	GTGGCTAGGAGCCATGGGAGGAAAGACAGCAGCAAGGGAGATATTTTCCACCGGTGACT	80223
Alignment Scores:		Qy	1080	-----	1080
Pred. No.:	1.14e-60	Db	80222	AAGAGGACGGTCGTTAGAGGGAGGGGACGGCAGTCTTGATAGTTTCAGGAAAGCCACCTA	80163
Score:	3445.00	Qy	1080	-----	1080
Percent Similarity:	18.65%	Db	80162	GGTTTCTTATTCTTGATGTCCTCACCACAGTAAAGTAAATCCAGCCAGGGGCCACCGG	80103
Best Local Similarity:	17.71%	Qy	1080	-----	1080
Query Match:	26.07%	Db	80102	CAATGGCAATGTACAGTCAGTGGTGGCTTAGTACAGTCGACCTGAGGGACATGTGCTACT	80043
DB:	2	Qy	1080	-----	1080
US-09-522-753-5 (1-2517) x AC136560 (1-218129)		Db	1080	-----	1080
Qy	905 ThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAsp	Db	79922	AAATTACTACCCAGCAGCCAGGGATCCCAGTCACGTGCCACTGAAATCCATGCCGA	79863
Db	80965 AGCACACGAGTGGCGTGGCCCACTCTCGAGTAGTGCCTGAGCCACATCCCTT	Qy	1080	-----	1080
Qy	925 GluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeu	Db	79862	GGTGACTCTCCCTAAGCGGGGGGGGGGATTTGTAGGCTCACAGTTAGGAGAGGT	79803
Db	80905 GCCATAGACCACAGTCACGGAGGTCAGCAA	Qy	1080	-----	1080
Qy	945 LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys	Db	79802	TGGCCTTACTGACTCTGGGAAAGATCCCTTTCGGACCCTGGGAGGAGAACAGAATCTAGGGA	79743
Db	80860 CTTGTCCT	Qy	1080	-----	1080
Qy	965 GlnLeuLysGlnArgAlaAla	Db	79742	GGGGGTAGAAGTGGGTGGCTTAGAGCCCTGAGTGTGGATCAAGAGGTGATGTGCA	79683
Db	80821 ACCGTGAAGCAAGATTAGCCCTCTGACACATGCCCTCTGTCATGCTCTGCTGTGGC	Qy	1080	-----	1080
Qy	980 LysValHisGluProProArg	Db	79682	GCCGTGGAGGATCCAGACTGGGGTCTTGTGTGGAGGGCTCAGGTCAATTGAAGCAGAGGTA	79623
Db	80761 TCAGCCACAGGACGCCCCACACCAAGCAGTTATCATGGGAACATAGCACACTTGTCTG	Qy	1081	-----	1086
Qy	994 ProAlaProAlaProProProProGlnAsnLeuGlnProGluSerAspAlaProGln	Db	79622	TATAGAGAGCCCCCACAGTATACACTTCTCTTCTTCTTGGAGGACACCTCTGCTCT	79563
Db	80701 ACATGTCCAATGAGCCCTCATGTGAGATGATGTCATAGTGCACCATGGAGATGTCAG	Qy	1086	uGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnProPr	1106
Qy	1014 GlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGlu	Db	79562	GGGCTTCCACGATAGTGCCTGGCGGGTCTCTGCCACGTCCTCCCCC-ATCTTAACCCCC	79506
Db	80641 AGCCCTCCCTGACTGCCAACTT-AATCTCTTTCAGCAGAGAAAGTCTTCTTTCCC	Qy	1106	oProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSe	1126
Qy	1034 AlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu	Db	79505	ACCCCTCATCTCTCTGCCAAGACACCTGTGTACTTTGAGAGGCAGCTGGTACCATCTC	79446
Db	80582 GCCTTCCAGCTGAGGGCCGAAGTCTCTACTGAGCCCCACGCTGGACATCGGGTCTG	Qy	1126	r-----	1126
Qy	1054 ProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspProSer	Db	79445	CCAGGTGAGTACTCTTTGGGTGGGTCTCCCTTCACGGAGGCCTTTCTGAGCGGTGGGA	79386
Db	80522 CTTTTCCTTCCTCTCTCGGAAGTATCAAGACTTCCCCACACGCTGCAGACCCCTCT	Qy	1126	-----	1126
Qy	1074 AlaPheSerTyrAlaProPro			-----	1080

Db	79385	TGTTGCCAGGGTGGTCTGGCTGGACTCAGCCCTGCTGTGTGGGCTACGTGTGTCA	79326
Qy	1126	-----	1126
Db	79325	GGCGTGTGTCCAGTGCCTTAGCCAGGTGACTCACTGTCTGTAGGCTTCAGTTTCATCCT	79266
Qy	1126	-----	1126
Db	79265	CTGTTCACTGGAGGGTGGGAGCTGCAGCTGCTCAGGCTGTAGAGTGTCACTGTCTGTC	79206
Qy	1126	-----	1126
Db	79205	GTAAGCCCTCGTCACTCTGCTGATCTTACTGACAGTCAACCACTCACCATGTTTGCA	79146
Qy	1126	-----	1126
Db	79145	AAGCAGTCACACTGTCTCTTTCTGTAGACATGGGCTTGGGCTGTGTGGCCTGGCCACA	79086
Qy	1126	-----	1126
Db	79085	GTCTCCAGTCGAGGGCGGCAGCCCGTCTCTCACTTGGGTATCTTAAAGTCTTAAGCA	79026
Qy	1126	-----	1126
Db	79025	AGGCTCTTCTGTGTCTGAATCTGGTCTTCCCGTGGCTGAGTCTGGTGTGATCTATCTC	78966
Qy	1127	-----GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAl	1143
Db	78965	TTGCGCCCTAGCAGGAATGTCACTGTCAGCTTCGGTGTCCCTCACTCAGAGCATGCCAAGGC	78906
Qy	1143	aProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys	1160
Db	78905	CCCCATGGGCCCTCTCAACCATGGGACTGCTCTCGCTGTGACCCCTAAGAAAGTGGGTGA	78846
Qy	1160	-----	1160
Db	78845	GTGCCCATGACCTTTTGGTTCTTATGCCCCGGGGGAGTCTTGGCCCTGGGAACTTAGAG	78786
Qy	1161	-----LeuAlaProPheSerGlyV	1167
Db	78785	GGGGGGGAGCCTGAGACTCTGTGTACCTCTCGCCCACTAGCCCCCTTCAGTGGAG	78726
Qy	1167	allYsGlnGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValP	1187
Db	78725	TGAAGCAGGAGCATGTTGTCCCTCGGGTCTCAGGCTGGGCCACCTGAAAGTCTGGGAGTGC	78666
Qy	1187	roThrAlaGlnGluAlaSerValLeuArg	1196
Db	78665	CTACTGCTCAGGAGACCTCTGTGCTAAGAGGTAAAGTCTGGCCCAAGTCACCTTCAGTCTTA	78606
Qy	1196	-----	1196
Db	78605	GAGAGCTGAAGCCCACTTGACAGTGTCTCTTACCCTTGACTCCACAGGAGACCGCT	78546
Qy	1197	-----GlyThrAlaLeuG	1201
Db	78545	AGGTGGCTTTTAGACTTGGTCTAAGAGCTGTGCTTATGCTTCTTAGGGACAGCACTGG	78486
Qy	1201	lySerValProGlyGlySerIleThrlyGlyIleProSerThrArgVal	1217
Db	78485	GCTCCACTGCCAGTGAAGATCATCCAAGGGCTCTCCCAAGTACCCCGACTGCAGATGGCC	78426
Qy	1217	-----	1217
Db	78425	CCAGTACCAGGCTCTATCACCCACGTAGTGTCTGGGGGAGGGTGGATGGGCAGGCA	78366
Qy	1217	-----	1217
Db	78365	TCCTAGGAGGTTGGGATGGATGAGAGGACCTCCGGGGAGACGAGAGATCTTGAGCTGA	78306
Qy	1217	-----	1217
Db	78305	ATGGTTGGTCACTCGGTGTGGGTCTAGGCTCAGGCTGATGAAGAAACAGGTGTCTATGT	78246

Qy	1217	-----	1217
Db	78245	GTCTCCAGCAGTTGGGTGTGGCAGAGACCAACACAGAAAGGATAGCCAGGCCCTTAACC	78186
Qy	1217	-----	1217
Db	78185	AGACTGTCTCAGAGCCATAGATCATAAAGTGTCTGGACTGAGAAAGTTTACCAGGTGACCCC	78126
Qy	1217	-----	1217
Db	78125	AGTGACATGTTGCCATGGCCAGAAAGACAGAGGGGACAGAGGTGGCAGCTCTTGTA	78066
Qy	1217	-----	1217
Db	78065	CAAAATCACCAGTCGGTAGGCATGAGCCAGTGGTGGGTACCAAGGCTTAGTACCCACGGTGT	78006
Qy	1217	-----	1217
Db	78005	GTGGCCAAGCATACAGCCAGCGAACTTCCAGTACCGGACCAACCACACAGAGTCTCTT	77946
Qy	1217	-----	1217
Db	77945	TTGACTCACTGTAGGGGACAAATGTCTCATTTATGCCCCAGGCTTCCCTCTGAAGTCTCTGG	77886
Qy	1217	-----	1217
Db	77885	AAATCTTGGGCTTTCAAAGGCAAGCTCAGAAGCTGGAGGGGAAATTCAGTCATGTCTCT	77826
Qy	1217	-----	1217
Db	77825	TCTGGCAGTCCAGTAGCCCTCCTCGACTCAGTCTTCCCTGTGACAGAAATGTTCTTTCC	77766
Qy	1217	-----	1217
Db	77765	AGGATAGGTACTCCAAGGAGCTCTTAGAGTGCCAGGTTTGACACAGAGCTTCCAGGGC	77706
Qy	1217	-----	1217
Db	77705	TCAGGGATGCCGCTCACAGGCTGAGNATCCCTTCCCTCGTGTAGGGCTTGTGGAGAGT	77646
Qy	1217	-----	1217
Db	77645	GAGGTCAAGCCTCTGGGGCTATCTGTGTGGATGATGACCCAGGCAATTGATGTGACACG	77586
Qy	1217	-----	1217
Db	77585	GCTAACAGGTGGAGACACACATCCCACTGTGCTGACCGCCCTGGAAACCAAGGCCAC	77526
Qy	1217	-----	1217
Db	77525	AGAGCCCATGTCCCACTCAGCTCCCAAGGATTACACTCAGGCTTGGAGGGGCTGCGGG	77466
Qy	1217	-----	1217
Db	77465	TGGGCCCCCTGTGCTCCTGTATTTCTTTGTTTGTCCAGTGAGTTATACGGAGCAGAAC	77406
Qy	1217	-----	1217
Db	77405	CATGGCTGGCTGTATTTCCGTTGTGACAGGCCCTCTAAGCTTGGAGACTCAAAACATC	77346
Qy	1217	-----	1217
Db	77345	AATACCCCCCTTTTCATCTGGAGCTGCTGACCTTTCTCACCGTATATACCTGGCCTAGCA	77286
Qy	1217	-----	1217
Db	77285	GCAGTACCAGCCACTTGAGAGGCAATTTCCCGACACACCCAGAAATCACCCCGCATGCC	77226
Qy	1217	-----	1217
Db	77225	CCACGGGGACCAACCAACAGAACAGTCCCAACGGGAGACAAACATCACTGAATCTGTCC	77166

Db 75005 CCGACACGCTGAGGCTTCACATGTCTCTTGAGACCGGTGCTCACACATCCCTTGTGTACA 74946
Qy 1281 ----- 1281
Db 74945 TGCAGACCTGCAGCGCTTGTGGGGTGTCTCTCGGGACTCAGACCGTGTGACATCTTC 74886
Qy 1281 ----- 1281
Db 74885 GTTAGTGTGCACATACACCCATGGGCACAGACACATACACCCCTCCTACACAGCAGTGG 74826
Qy 1281 ----- 1281
Db 74825 CAGGCATGCCCTGGCAGCTTTCTGCCAAGCAGAGTTCACAGTTGAGCTGGTTGAGGGTCCC 74766
Qy 1281 ----- 1281
Db 74765 ATCTGTGGGGCAGGCGCTCTCTTCAGCTCTGCTGTTTGCCTAATCCATGCCCTGTGGT 74706
Qy 1282 ----- GlyGlyMetSer 1285
Db 74705 GTCTGCACCTTGTCTGCCACCCACACCATGTGCTTTTCTGGCCCTCAGGTGTATGTG 74646
Qy 1286 ValThrGlnCysSerIysGluAspGlyArgSerSerGlyProProHisGluThrAla 1305
Db 74645 GTGTCACAGTGTCTCAAGGAGGATGGGAGGAGCTCAGGCCACCCACGAGACTGCC 74586
Qy 1306 AlaProIysArgThrTyAspMetMetGluGlyArgValGlyArgAlaIleSerSerAla 1325
Db 74585 GCCCTTAAACGCCTATGACATGTGAGGGCCGTGTGGCAGGACCATCACCTCAGCC 74526
Qy 1326 SerIleGluGlyLeu ----- 1330
Db 74525 AGCATCAGGGTAA-GTGTCTCACTGTAGTAGTTACCGGGGTCAAGCACCATCCAGGA 74467
Qy 1330 ----- 1330
Db 74466 GACCCAGGTACCCAGTAAAGATGSAAGGAGGTGGGCCCTTGACATGCCTGGAGGG 74407
Qy 1330 ----- 1330
Db 74406 AGCCAGTCAGTAGTTGGCTCAATATTAAACAGTAGAANAAGTAGGCACACCCGGTGCA 74347
Qy 1330 ----- 1330
Db 74346 AAATGCAGCCCGGAAACGGGCACCTTGTCTGGGTGTTAAGTCACAAACAGAGACTG 74287
Qy 1330 ----- 1330
Db 74286 TTGGTGTAGAGCCAGCCCTCTCGSACACATACGGTGACACTATCCCTGCTCTCCCA 74227
Qy 1331 ----- MetGlyArgAlaIleProProGluArgHisSerProHisGluLeuLysGluGln 1348
Db 74226 GGACTCATGGCCGTGCATC---CCTGAGCAGCACAGCCCC---CACCTCAAGAGCAAA 74173
Qy 1349 HisHisIleArgGlySerIleThrGln----- 1357
Db 74172 CATCACATCCGAGGCTCCATCAGCAAGGTATATCTCTTACGGTGTGTAATTCACCT 74113
Qy 1357 ----- 1357
Db 74112 CCAGCCAAAACCCAGCTAAGGGGTGGAAACCGAGGGTGAACCTTCCCGTCCGTCTTAT 74053
Qy 1357 ----- 1357
Db 74052 AGGTGAGTGACTACTCTGTACTTGTGGCCTCATTTCCAAAGCCCATGTGATGCCCTTGCT 73993
Qy 1357 ----- 1357
Db 73992 AGTGATTGAGAGGGGTCCAGATTGAATGTGTCTTTAGCTCATGCCCTTGTGGTAAACTG 73933
Qy 1357 ----- 1357
Db 73932 AGCAATGTGTTTCACTTAGCGTCTCTCAGTCAATGAGAAACAGGCTTAGAAAAAGTCA 73873

Qy 1357 ----- 1357
Db 73872 GTTTGCATACCCAAAGTCCACAGGAAGATTGTGCTGACCCAGGCTGAGCCAGGCAAG 73813
Qy 1358 ----- GlyIleProArgSerTyValGluAlaG1 1367
Db 73812 GCAGTGGAGAGCCAAATGATGTTGCCCGCAGGATCCCGAGGTTCATATGTGGAGGCCCA 73753
Qy 1367 nGluAspTyLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProPr 1387
Db 73752 GGAGGACTACTTACGTAGGAGGCGCAAGCTCTTGAAGCGAGAGGACGCCACCGCCCC 73693
Qy 1387 oProProSerArgAspLeuThrGluAlaIleTyLysThrGln-----AlaLeuGlyPr 1404
Db 73692 ACCGCCACCTCGGACCTGACTGAGACCTACAAGACCGGCCCTCGACCCCTCTGGGTCC 73633
Qy 1404 oLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSe 1424
Db 73632 CCTGAAGCTGAAGCCAACTCATGAGGTGTGTGAGCGACTGTGAAGGAGGCGGCCGCTC 73573
Qy 1424 rIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProAr 1444
Db 73572 TATCCATGAGATCCACGAGAGGAGCTGCGCCGCGACCTGCGAGCTGCCCTTGGCACCACG 73513
Qy 1444 gProLeuLysGluGlySerIleThr----- 1452
Db 73512 GCCTCTGAAGGAGGGTTCATCAC-CCAGGTACGAGGAGCACAAAGGAGGCGCGGGCT 73454
Qy 1452 ----- 1452
Db 73453 TGGCAAGCCTGAGGGAGTTCTAATGTGCCCTCATTTGCATTTGGGTTTATAAACTAAGGCC 73394
Qy 1452 ----- 1452
Db 73393 CAGAGAGTCAGCGGCTTGTCTTGGTCACTGACAGGCTTGTGAAACCAAAGTCTTTGG 73334
Qy 1452 ----- 1452
Db 73333 TCACTTGGCTGCTCCCCACATGTCTACCTGCATGTTGTGTCAAGACTAGAGCCACTGT 73274
Qy 1452 ----- 1452
Db 73273 CCCCTACCTTGGTTCATGGCTTGATGGAAAAAGAAAGACAAACAGCCAGAACGACCTA 73214
Qy 1452 ----- 1452
Db 73213 GGGAAGGCTGTGTCACTGAGCCAGGGACACACAGTGATGGACCAAGATGATTCTAGG 73154
Qy 1452 ----- 1452
Db 73153 GTATTCTTTTGAAAACTTAGTGTGTCTCTCATTTGGGCCCCAGCACTTGGGAGCGGAG 73094
Qy 1452 ----- 1452
Db 73093 TCAGGAAGATTGTAAATTTGAGGCTAGCTAGGCTACAAAGCTGTACTCTGGCTTACTGC 73034
Qy 1452 ----- 1452
Db 73033 CGATCACCAAAAGTTTACAAATTTAATGACAACAACAATTGAGTGAAATTGAGATCTGG 72974
Qy 1452 ----- 1452
Db 72973 GACCTAGCTGTCTACTGTCCCAGGCTGGATCCCTGTGGGATCCCTCTCCAAGATGTCC 72914
Qy 1452 ----- 1452
Db 72913 AATAAATTCGGTCTCCGTGGTGTCTCTGAGAGCAACAGGAAATGTGTAACTCCAGCCT 72854
Qy 1452 ----- 1452
Db 72853 GGGACCCGGAGCCAGAGCTGGGTGACAGCAGATGCATCTGTATATGGAGTCCATAGTGCT 72794

QY 1452 ----- 1452
Db 72793 GCAGGATGTCGGGGGCCACACTGGGTGAGGAAGGTGGAATTGACCTGTAGTTAGCCACGC 72734
QY 1452 ----- 1452
Db 72733 TCTCGTACCCTCCACAGTGGGGGTCTCTGGAGCTATAGCCAGGCTGAAGCGAGCAGG 72674
QY 1452 ----- 1452
Db 72673 TATCTAGCCTCTGACGCTCTGCTTGTCTCAAGGGGTTAGAGATGCCCATCTCTGTGCAGAC 72614
QY 1453 -----GlnGlyTh 1455
Db 72613 ATACAGGAGAACAGCATCTGGCCAGCCACCATACAGACCTGGACCTCTCCCCAGGGGAC 72554
QY 1455 rProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSe 1475
Db 72553 CCCACTCAAGTAGCACTCCGGGGTGCCTCCAGTGGCTCCAGAAACACAGAGCTACGCTC 72494
QY 1475 rLeuileGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAs 1495
Db 72493 CATCATCGGAGCCCGCGCGGCTTTTCTGCTGCCCTGCACCCACTGGACATAATGGCTGA 72434
QY 1495 pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAl 1515
Db 72433 TGCCCCGGGCACTAGAGCGCGCTGCTATGAGAGAGTCTGAGAGCGGTCAGGGACCG 72374
QY 1515 aSerSerSerGlyGlySerileAlaArgGlyAlaProValileValProGluLeuGlyLy 1535
Db 72373 CAGTGGTGCAGGGGCTCCATCACACGCGGGGCCAGTTGTGCTGCTGAGCTAGGCAA 72314
QY 1535 sProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr 1555
Db 72313 GCCCCGGAAGGCCACTGGCTACGAGAGCCACGCGGGCACCTTACCAGCCACCTGCCC 72254
QY 1555 oArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSe 1575
Db 72253 CCGAGGCTCACTGTGACTACAGGGAGCCACCGCCAGCGCTTCAGGAAGGTGAGGC-TC 72195
QY 1575 r----- 1575
Db 72194 CCCTGGGTGCCAGCTGTGGGGGGGAGACAGCTCATTTTCCACGCGCAGTCTGTCTGCTG 72135
QY 1575 ----- 1575
Db 72134 CCCTGCCCTCCACTGCATACAGAGCCCCCGTCTCCCTTACCCTTACCAGCATGA 72075
QY 1575 ----- 1575
Db 72074 TAATCAAAATCAGGGCTGAGGAGGGCTCAGTAGCTAAAGTCTTGTGTACCAGCATGA 72015
QY 1575 ----- 1575
Db 72014 GACTCAGAGTTCAAATCCCCGGTACCCACATAAGCCAGGCACACTCATACACTCTGTAA 71955
QY 1575 ----- 1575
Db 71954 TTCTCATGCTTCACAGTGAGACAAAGGCTAAGACAGAAATCTCCTAAAGGTTATAGGCCA 71895
QY 1575 ----- 1575
Db 71894 TGAGTCCGGGTCTCAAAATCGGGAAGTGGATGACACCTCTCCCCCCCCCTCTATT 71835
QY 1575 ----- 1575
Db 71834 TTATACACGATACGACACTCATATACAGCACACTCGCGCTACATTCTCTGTGCAC 71775
QY 1575 ----- 1575
Db 71774 ACATGTACACAAAAACAATAGTTAAATGAAGCAGTTCCCACTCTGTAGGCTTCACCT 71715
QY 1575 ----- 1575

Db 71714 CTCCAGTAGCGCTGCTGGTCTGAAGCACCTGCTGTGACGTGACCGCGCATCTCTACTCT 71655
QY 1575 ----- 1575
Db 71654 GAATCCGCTTCTGATTCTCTCGCATTTCTGACCTCCGCTCTTTCTCCGATCTCCACAT 71595
QY 1575 ----- 1575
Db 71594 GTGAGCCTTTCCGTGTGTGCTGCGCTCTATGCCGAAGGCTCCACCTTTGGTTTGTCTCT 71535
QY 1575 ----- 1575
Db 71534 AGCCAAACCACTGTATATACCCACGTCACCAAGATGTCAAGGCGCTGGAGAGCTATAGGCACC 71475
QY 1575 ----- 1575
Db 71474 TCCAACTAGTTCCAGCCACATCCGGCGCGGGATGTGTTAGGTTGTACAGACAGTGTCTCTCT 71415
QY 1575 ----- 1575
Db 71414 AATTCAATTATGAGGGAGCCCTTCCAGAGCACTGAGATCTCTCAACATCTCTCTCTCT 71355
QY 1575 ----- 1575
Db 71354 CAGGCGCGTCACTTCTTACCTTTGGGTTTCATCTCTATCTGGAACAATGGCTCTTTGGAG 71295
QY 1575 ----- 1575
Db 71294 CTCAGGGTTCTTGGCTGTCTTCTCTCAGGGTGTCCCTCTCTCTCATGGGAATGCC 71235
QY 1575 ----- 1575
Db 71234 AGCCCATGATTAAAGTTCAGACTGGAATGAATAGGCAATGGTGACTGGGAGTAAATGAA 71175
QY 1575 ----- 1575
Db 71174 TGCTTTCAAAAGGAGCAGGAGCAGAGAGACAAGAGGCTCTAGGGTCTGTGACAGCTCAG 71115
QY 1576 -----Se 1576
Db 71114 CTGTGTGGCCCGCTTGAGGCTGAGGCCACTTTGACCTGTCCCCACAGGCAGCTCTCTC 71055
QY 1576 rSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPr 1596
Db 71054 CAGCAAGGGCTCCAGGACCGGAAAGTGAACCTCGACACACCCCGGAGATTGCCAAGTCCCC 70995
QY 1596 oHisSerThrValProGluHisHisProHisProHisProHisProHisProHisProHis 1616
Db 70994 ACATAGCACTGTGCCGAGCACCCACCCCATCTCCCTCTAGAGCACTGTCTCCG 70935
QY 1616 gGlyValSerGlyValAspLeuTyrArgSerHisHisProLeuAlaPheAspProThrSe 1636
Db 70934 GGGTGTGACTGTGTGTGAGCTGTACCTGTGGCCACATCCCATTTGGCCCTTTGACCCCACTC 70875
QY 1636 rIleProArgGlyIleProLeuAsp----- 1644
Db 70874 CATACCCGAGGATCCCTCTGGAAGCAGGTGGTGCCTGGGCTGTACATGTGGGGCTT 70815
QY 1644 ----- 1644
Db 70814 ATTGATTGTCTTCTTGTGGTGGACACTTGGGTATGGCTTAGTGTGCTGCATCACCAGCT 70755
QY 1644 ----- 1644
Db 70754 GCTTACCCAGGCGAGTCTTAGGATTTGGGCTTTTGTGGTGTCTTCTCAGCTAGGACT 70695
QY 1644 ----- 1644
Db 70694 CATCTGTCTTGCAGTGCAGTAAATAGGGGCTGCAGTGGCCATGATGTGGAACC 70635
QY 1644 ----- 1644

Db	70634	TTTTAAAGGCTCTTTGTGAAGGTGTAGCACTGTATTTTCGATGCAATATATATGGGGG	70575
Qy	1644	-----	1644
Db	70574	GCACCTTCGAAACATTAAAGCAGGCCCCCACCACATGATGAGGAGCGAGTCTTAAGCTC	70515
Qy	1644	-----	1644
Db	70514	TGTGTGTGCATTGGGGAGGCATGTGGGTGTACAGGAGAGAAAGCTGTTGTGCAGAAAGT	70455
Qy	1645	-----	1645
Db	70454	GAGGAACGTGAAGATAAGTCATNGGGCCACCTGACCCCTCCATGCCCTCCCTCCCGCCAG	70395
Qy	1645	aAlaAlaLayrYrLeuProArqHisLeuAlaProAsnProThrYrProHisLeuTy	1665
Db	70394	AGCGCAGCCTACTACCTGCCCCGGCACTTGGCCCCCAGCCCCACCTACCCACACCTGTA	70335
Qy	1665	rProProTyrieullelAargGlyYrProAspThrAlaAlaLeuGluAsnArgGlnThrIl	1685
Db	70334	CCCCCTTACCTCATCGCGGCTACCTGTGACACCGCGGCGCTGAGAAACCGCCAGACCAT	70275
Qy	1685	eileAsnAspYrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGl	1705
Db	70274	CATCAATGACTACATCACTCGCAGCAGATGCACCAACGCGACCTCTGCCATGGCCCA	70215
Qy	1705	nArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyAl	1725
Db	70214	CGCGCCGATATGCTGAGGGGTCTGTACCAGAGAGAGTCTCTGCTGGCCCTCAATATGC	70155
Qy	1725	aAlaGlyProArg	1729
Db	70154	CGCTGCCCGCGTGTGATGTTGCTGGGCTGTGCCCATCAGCAGAGACTTGTCTCTGCGC	70095
Qy	1729	-----	1729
Db	70094	TGTGGTGTGCAAAAGAGATGTCACGTGTGCACTGCAGCTGGCGGAGGGGCACAGATT	70035
Qy	1729	-----	1729
Db	70034	GGCTGGGAGNACTAGGTGGTGATGCTGGCTCTTGGGGGAAGGCAGCGGGCCACTT	69975
Qy	1729	-----	1729
Db	69974	GTATCTGTTACATCTAGAACTGTGCTTAGCAGACACTCTTGCCACCATGCACCGTGGGT	69915
Qy	1729	-----	1729
Db	69914	CTTTGCGTGTGTTCTGAAACCTTGGTATGGGCATCTTCTGTATCCAAGAGATCAAAA	69855
Qy	1729	-----	1729
Db	69854	TACAGGCTGGATGTGGCATTTGCACGCCCTATATCTTAAGTAGAGGAAGGTGCAAGGCCAG	69795
Qy	1729	-----	1729
Db	69794	CCTCAGCTACATAAAATTTTGAGGCCGCTGGGCTATATGAGTCTGCCTCTATACAAATG	69735
Qy	1729	-----	1729
Db	69734	ACTGAAAGAAAGGAATGAGAGCATGCGGCTTGTTCAGAACACTGACTTCTGTGAAACAGC	69675
Qy	1729	-----	1729
Db	69674	CTGGGTCACTTTGCTCGACTGTTTCAGAGATAACCTTGTGGCCCCAGTTAGTGATACGGA	69615
Qy	1729	-----	1729
Db	69614	AGCCACTCGGGGTTTTAGAGGATGGAGCCCAATGCAAGGAGCTGATACTGTGGGAAG	69555
Qy	1729	-----	1729
Db	69554	GCAGTGTGGGGTCAAAAAGATGTTGTTTGAGAGACACTGCAGGTCTATAGCCTTTGAGAAA	69495

Db	66297	ACCACCTGCACCTTCGTTTTGGGAGTGGGATGACGGGTCTACAGTTGACTGTGCCAC	66238
Qy	2039	-----	2039
Db	66237	CACGTACATGGAGAGTTAAGCAGGAGTAGGTGGGGGAGGCTGGTACCCACC	66178
Qy	2039	-----	2039
Db	66177	TCCTCTCCACACACAGCTACAGCCCTTAAGGGAATGAAGCCTCCCGCTGGATCAG	66118
Qy	2039	-----	2039
Db	66117	AGACAGTGGGACTGGGCCAAGCTGGACCTAGGACCCATTGTCACTCAGCTTGAGCCT	66058
Qy	2039	-----	2039
Db	66057	CACACACACACACCTGACCCCATGACATTCACACCTCCCTCCAGTTTCTCCACACC	65998
Qy	2039	-----	2039
Db	65997	ACCTCCAGTCTCCCTATCTGAGCCCCACACGCCCAAGTCAGACTCCCATATCCTTT	65938
Qy	2039	-----	2039
Db	65937	GGGTGTCTGAGTCTCATAAATGGAGCTCCCTTCACTTGCAGAGCTTCTCTTGSCC	65878
Qy	2039	-----	2039
Db	65877	ACCCCTAGTTCGACCTGATTTTCTTCCATCTTGAGGCCACAGTGTGGAGAAATGGCTCC	65818
Qy	2040	-----SerLeuGlyTyrHis---Glys 2046	
Db	65817	TCCCAAGCCCCACCACTTCTGACCCCTAGCTACCTGCTCTCA-GGTTACACAGTGGAG	65759
Qy	2046	erSerTyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrH	2066
Db	65758	CTGGCTACAGCCCTGTAGTGGGTAGAGCCATCAGCCCGTGTGAGCTCGCCAGCTGACCC	65699
Qy	2066	isAplysGlyLeuProLysHisLeuGluLeuAplysSerHisLeuGluGlyLul	2086
Db	65698	ACGACAAAGGGGCTCTCCAGCCTCTGAAGAGCTAGAGAGAGCCACTTGGAGGGGAGC	65639
Qy	2086	euArgProLysGlnProGly-----	2092
Db	65638	TGGGCACAGCAGCAGCAGGTGACACCCGGGAGAGTACAGGTGGGAAACTGAGGCAGG	65579
Qy	2093	-----ProValLysL 2096	
Db	65578	CTCCATGTTCTCCTCAGGCCCTGACAAAGCTCTGTTGTCTCTCAGGCCCATG-AAAGC	65520
Qy	2096	euGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerS	2116
Db	65519	TCAGTGGGAGGCTGCCCACTCCCACTTCGGGCCACTTGGCGAGAGCCAGCCCTCAT	65460
Qy	2116	erSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuA	2136
Db	65459	CCAGCCCACTCTCCAGCTGCCCGCCAGCATCAAGGTCAAGGTCCACAGAGGGTGTACCTTG	65400
Qy	2136	laGlnHisIleSer-----	2140
Db	65399	CTCAGCACATCAGTGAAGTGCCCATCTCTGGCCCTCCATTTGGGAAGGCAGGTGGCTA	65340
Qy	2140	-----	2140
Db	65339	TGACAGCCATATCCACAGCCAGACAGATGCTAACTGCTTAATCTGCCACCAAGCTTAG	65280
Qy	2140	-----	2140
Db	65279	TCCTTTGGGGGTTCATCATAGGTCTAACAGGACACTTACCAGGAGACCAAGAGCCAG	65220
Qy	2141	-----GluVal 2142	
Db	65219	ATCAGGGTTCTGCCCTACTCTGAAGTCTGAAAGTCTGAACTGCCTCTCCTCGCAGGAGGT	65160
Qy	2143	IleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaPro	2162
Db	65159	ATTACACAGGACTACACACGCCACCCACAGAGCTCAGTGGCCCCCTTCCCGCCCT	65100
Qy	2163	LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp	2182
Db	65099	CTCTACTCTCTCCCTGGAGCCAGCTGCCCTGTCTGGATCTCCGCCCTCCACCCAGTGAC	65040
Qy	2183	LeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGly	2202
Db	65039	CTCTACTCCACCCCGGACCATGGCACCCAGCCCGGGATCCCCCAGTGAAGGG	64980
Qy	2203	GlyLys-----	2204
Db	64979	GGCAAGAGGTGAGGAAGTACCCATGTCTGCATCAGGTGCTGGGAACAGTGGGACAGGGT	64920
Qy	2204	-----	2204
Db	64919	GAGCTTTGAGTGTGGGCTGATGGGAAACTGAGGTGCAGGGATACAAAGCCTCACTACG	64860
Qy	2204	-----	2204
Db	64859	CTCCAGAACAGACACAAAGCATCTAATCAAGAGGTGGCTCACCAAGGTCCACGTG	64800
Qy	2204	-----	2204
Db	64799	CCAAATTGAATGTCTATATATACCACTCAGTAGGGGTGAGTGTCCCATTTCAATT	64740
Qy	2204	-----	2204
Db	64739	GCAGATAGGAAACTGAGGCATAGAGAGGTGAGTTACTATCCCGTCCCATGCACAAAC	64680
Qy	2204	-----	2204
Db	64679	ACCCTGCACAGATGTAGACAGGTATCTCTGGGTAAACAGCAGGTCTTAGCATGGGGCCT	64620
Qy	2204	-----	2204
Db	64619	CATCAGGGCTTACAACTGCTCCAGCACCCAGCAGCACTGTGTGGTGGCATACCTC	64560
Qy	2204	-----	2204
Db	64559	TGCACAGGCTGAGCAAGACGACTGTGAATGTCAGGGCAGCCTGGGTGTACCAAGAC	64500
Qy	2204	-----	2204
Db	64499	CTGACTCAAAATAAAACCAAGCCCATGAGTGTGGGTGAGTTCGAAAGGTCTCAGAAG	64440
Qy	2204	-----	2204
Db	64439	CCCCCTGGGTCTCCATACCTCAGATGCAGAGTCTGCCATTTCACAGGGAACACGTGAG	64380
Qy	2204	-----	2204
Db	64379	CCCCACTGTGGCAGTAGGGGTCTCTTCATGAGTAGAAACCCCACTCTCAGCTGGAT	64320
Qy	2205	-----ArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAsp 2220	
Db	64319	GCTGTCTCCATTAGGTCTCCAGAACCCAGCAACA---GTCCCTGAGTAGCAGCGAGAT	64263
Qy	2221	GlyIleGluProValSerProProGluGlyWetThrGluProGlyHisSerArgSerAla	2240
Db	64262	GCTATTGAACCTGTGTCCCCACCGAGGGCATGACTGAGCCAGGACATGCACGGGCACT	64203
Qy	2241	ValTyrProLeuLeuTyrArgAspGlyGluThrGlu-----	2253
Db	64202	GCCTACCCACTGTGTACCGGACGGGAGCAGGGGAGCNCAGGTAGTATGGCACA	64143
Qy	2253	-----	2253
Db	64142	TCCCCGTTAGAGGTGGTGGTGTGGTTTCCCATACACCCGTGGCCTCAGCCCTCTAGCG	64083

QY	2253	-----	2253		
Db	64082	TGGAGGTGCGTCAATCTCTTACAGCTCAGCGGCTGATGACTTTTCAGGGGCTTTCAGTGG	64023		
QY	2253	-----	2253		
Db	64022	TGGCACCTTGCTGCTGGCATCTCAGCCCGAGTCATGTCCTATTATTCATCTTGCTGCTGAT	63963		
QY	2253	-----	2253		
Db	63962	TGAGATGAACATAACCGGACACAGAGTGGGTAGGGTGGGCATGGACCTCAGGAGGGTA	63903		
QY	2253	-----	2253		
Db	63902	TGACCTATGGCGGCCAGTCCCTGGGAATCTTGTTGTCAGTATAGTCTTAAGAAGTTCT	63843		
QY	2253	-----	2253		
Db	63842	CAGGGGAATTCACAGTGTGCTGGAGGCTCACAACCGCCTAAGCCAGGGCTAAGTA	63783		
QY	2253	-----	2253		
Db	63782	CAGATCCACAGAGCCCTGGGATGGGAATGGTTGGTCTCTTAGCACCTCACTGTGGCTCAA	63723		
QY	2253	-----	2253		
Db	63722	GTGGTGAGCTAGCCCCCTGTTGCTAAACACAGTGCCAGCCTGTGTACCCCTGTCA	63663		
QY	2254	-----	2272		
Db	63662	ProSerArgMetGlySerIysSerProGlyAsnThrSerGlnProProAlaPhePh TGTCTCCCTGCAGGATGGGTTCTAAGTCTCCAGGCAACACAGCCAGCCGCGCAGCTTCTT	63603		
QY	2272	eSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLy CAGTAAGCTGACCGAGAGCACTCCGCCATGGTGAAGTGAAGAGCAGGAGATCAACAA	63543		
QY	2292	sLysLeuAsnThrHisAsnArgAsnGluProGluTyr-----	2304		
Db	63542	GAACTCAAACCCATAAACCGGAATGAGCTGTAATACAGTAAGGAGAGCAAGGCCAGA	63483		
QY	2304	-----	2304		
Db	63482	GCAGGCATGGGACAGAGTGTCTTAAAGCCTTCTCGAAGGGCTTGATGTCGTGCC	63423		
QY	2304	-----	2304		
Db	63422	AGTGGATACCGACGGCATATAGCACCTCAGTAATGCCCTAGGCACAGTAGGTGCTCAGCG	63363		
QY	2304	-----	2304		
Db	63362	AAACCTATAGACAGAGCCCTGTGTGGGCGAGGAGCACAGTGCACCTACTCCAGGTTCA	63303		
QY	2304	-----	2304		
Db	63302	TGCTGCAGGCATGGCGCCTTTCTCCACAGCTGATGTGGGTGGACAAGGACCTTTGTCCTT	63243		
QY	2304	-----	2304		
Db	63242	AGCTCAGGGTGCAGGGACCATATGCTTATGAGAAATGCACCTGGGAACAACACTTGAGGG	63183		
QY	2304	-----	2304		
Db	63182	CTTCGCACCTGTCCCTCCCACAGCAGCCCTCCCTAGGTTGGCTTGGACAAATCCCTTC	63123		
QY	2304	-----	2304		
Db	63122	TGGAGTTAGACTACAGAAGCCCTTGCAGGTTAGCACAGGCCCTAGGGTAGTGAACCTACTGC	63063		
QY	2304	-----	2304		
Db	63062	CAGGTAAACCTGGAGCATCAGATACCTAGTTGTGCAGACATTCAAACCCAGACCGGTACA	63003		
QY	2304	-----	2304		
Db	63002	GGGAGCTAACGGGTGCAGATCTCCAGAACTCTGAGCTCAGGCCTTACCCACCTTGTTA	62943		
QY	2305	-----	2321		
Db	62942	CCCTGCTAGATATCGGCAGCCTGGGACGGAAATCTTCAACATGCCGCCCATCACTGG	62883		
QY	2321	YThr-----	2322		
Db	62882	AGC-AGGTAAACCTCCCGTGTGTCGCATAGCAATCTCCCTCCCACTACTGACTTAG	62824		
QY	2322	-----	2322		
Db	62823	TGACAAAGAGCCTTAGCCAGTTCAAATTTGAGCCCTCTCCCTGGCCTTAGTGGACCCAGA	62764		
QY	2322	-----	2322		
Db	62763	ATGGCACAGAGCCAGCCCTCACTGCCTCGATCTCTCCCTCATAGAGGACGTCCAAAG	62704		
QY	2322	-----	2322		
Db	62703	TTTGAGATGAGGGTTTTCTCCGGTGGGTGTAGGTACACCTTGGTAATACTTGCCCAATT	62644		
QY	2322	-----	2322		
Db	62643	GCTGGGGCTGAGGACGTCACCCAGTGTGGGCCCTAGGAAATTTGGCCTTGGGGTCCAGAGC	62584		
QY	2322	-----	2322		
Db	62583	TTCAGGGTGCACAGGCATCTCTCACATACAATCTCCATTTGGGGTTCCAGCAGAGAGC	62524		
QY	2322	-----	2322		
Db	62523	CCTGGCCTTTTCTAAATCAGATGCAACCGGTAATGTGAGAGAAACATGTCACACTCAG	62464		
QY	2322	-----	2322		
Db	62463	GTGCAGTTGTGCAGAAATTTTCAGGCATTCAGTGTGGTCTTGGGACTCAGAACTCGG	62404		
QY	2322	-----	2322		
Db	62403	CTTGCCATGTTTANGATGAGTTGCCCTTGGGAAAGAGCTGTCATGGGAATTAATGGCT	62344		
QY	2322	-----	2322		
Db	62343	CAGGAACCTGAGCCGGGACTGTGACACTGGGCGAGCTGTGAGGACTTGCACTCCAGGC	62284		
QY	2322	-----	2322		
Db	62283	TGGAGGAACAGGCAAGGGTTCACTACCTCTTATGTGTGGAGCTTTGGGCCAGGGAAT	62224		
QY	2322	-----	2322		
Db	62223	AAGTGGCCAGATACACCTTTGGGAACATTTCCAGAGGACCAGAGACCTGCAGAGGGGG	62164		
QY	2322	-----	2322		
Db	62163	TGGGGTGGGATCCTGCGAGGGCTGTTCACAAACCCAGAAACATACACTGGGGACTTTTCTG	62104		
QY	2322	-----	2322		
Db	62103	AGCCTCGGGCAGAAGGAACTTTTGACAAGTTGAAGATGAATGTTTCTAGAATGTTCTAAA	62044		
QY	2322	-----	2322		
Db	62043	TCTTGATAGGTTCTAGACACCTGGGTGTGTGTGAGGTAGAGATTTCACAGGGTAATACA	61984		
QY	2322	-----	2322		
Db	61983	CGCACACACACACACAGCGTGTGCTCTCAAACACCCCCCATAAATGTTAGTCTCTCAA	61924		
QY	2322	-----	2322		

QY 2397 ----- 2397
Db 59703 TCACACCACTCCCCACATCCACCCCTTCTCCCTCGTGTGTGGACAGGATCAAGCTG 59644
QY 2397 ----- 2397
Db 59643 ACCAGGCCACCATGTGCGACCGGGGCTACTGTGGGAAGAGTCTAGTGTGGCATGACC 59584
QY 2397 ----- 2397
Db 59583 CTTTGAAGGCCAGGCAGCAGAGGCCAACTCCCGGTCTTCCAGTGTGCTGTACGACTCT 59524
QY 2397 ----- 2397
Db 59523 CTGTTCCCAAGCAACACAAAGCAAGAGGTGCTCTGCTGTGCCCTTCTAGGGTCAGGCCA 59464
QY 2397 ----- 2397
Db 59463 GCCTCAGGGTCTGGCCCTATCAGCAGAGGCTTTCAGTCTAGCACAAATGAATGCTGTCTT 59404
QY 2397 ----- 2397
Db 59403 GGGCAATGCTGTCTCCAGCCAAAGATCTGACACTCTCCTAGTGAGGGGTACAGCTGGTG 59344
QY 2397 ----- 2397
Db 59343 GGGCAGCTAAGGAAATAGTCAACCACTCTCTCATAGCAGATGTTTTCCTCTTTTTC 59284
QY 2397 ----- 2397
Db 59283 ACGCTAGGGTGGCTCTGGCCATCTGTTAGACCTCTTAGCCTCTGGCTCAGGGTTTCT 59224
QY 2397 ----- 2397
Db 59223 CACTAGTCTTCTGTTGGGGCCCTAGGAGCAGGTGGCTCCCCAGGAGAAATCGGCC 59164
QY 2397 ----- 2397
Db 59163 CCACTGCCCTCTCCGTCTGTGTGGTGACCCACAGTCAACACCATCTAACTCAGCC 59104
QY 2398 -----GlyGlyGlyGlyAlaLysValSerGlyArgProSe 2410
Db 59103 CTTTCCCACTTTGTCCTCGAGTGGTGGTGGAAAGCCAGGTTTCTGGCAGACCAG 59044
QY 2410 rSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProSerVa 2430
Db 59043 CAGCCGAAAGCCCAAGTCCGAGCCGCGCTGGCACCAGGAGACCCAGCACCTTCTGT 58984
QY 2430 lSerSerValHiserGlyAspCysAsnArgArgThrProLeuThrAsnArgValTr 2450
Db 58983 CTCCTCAGTACACTCGAGGGGCGACTGCAATCGCCGAACACCACTCACCACCCGTGTG 58924
QY 2450 pGluAspArgProSerSerAla----- 2457
Db 58923 GGAGGACCGCCCTCATCTCGAGTGGGTATCAGTAGGGGCGAGAGATGTAGACATGGGG 58864
QY 2457 ----- 2457
Db 58863 TTCCTGGGTACCCAGGGGAGGGCGTCCCTGGTCCCACTAGCAAGGAGGTGTAGG 58804
QY 2457 ----- 2457
Db 58803 AAGCCACACATTGAGAGGGTCTGTAGATCTCCTCAGGCGCTTCCAGTCAGGCTGGTGCAC 58744
QY 2457 ----- 2457
Db 58743 ACAGGGGACAGTGCCTGTCCAGCAGCAGTAGGATTTTGACAGGAAAGCAGGACGTAGA 58684
QY 2457 ----- 2457
Db 58683 AGGAAAGTTACACACACCCGAGTCAAGGTGGTAGGCTGTCTAGCACAGGCAAGGG 58624
QY 2457 ----- 2457

Db 58623 AGAAGGCCGCTTTTGGAGAACTCAAATGGTGTCTGGCTAAGAGGCCCGAGCTTC 58564
QY 2457 ----- 2457
Db 58563 CTGTGTGTAGAAATGGTCTGAGGTGGCTACTTGGAGAGCCCTCCACCATCTCCCA 58504
QY 2458 -----GlySerThrProPheProFyrAsnProLeuIleMetArgLeuGlnAlaG 2474
Db 58503 TCTTTCCACAGGGTCCAGCCATTCCCTACAACTTTTGATATGAGGCTACAGGCAG 58444
QY 2474 lyValMetAlaSerProProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyP 2494
Db 58443 GTGTATGGCTCCCAACCCACCTGGCTCCCGCAGGACGCGGCCCTAGTGGTC 58384
QY 2494 roHisAlaTrpAspGluPluProLysProLeuLeuCysSerGlnTyrGluThrLeuS 2514
Db 58383 CCACCATGCTGGGATGAGGAGCCCAAGCCACTGTGTGTTCACAGTACGAGACACTCT 58324
QY 2514 erAspSerGlu 2517
Db 58323 CGGACAGCGAG 58313
RESULT 17
AC121005 219339 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-322A11, WORKING DRAFT SEQUENCE.
DEFINITION AC121005
ACCESSION AC121005
VERSION 4 GI:25097679
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 219339)
Muzny, D., Marie, J., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Burch, P., Burrell, K., Calderon, E., Cardenas, J., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinu, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harveys, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.B., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhera, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweli, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,

Db	166306	TGGCAGTCGGCTCTTTGGCAGCTGGCGCAGGGCTGCATGCTGCTGGTGGTGAGATGAGA	166365
Qy	1080	-----	1080
Db	166366	CCTTGCTGAGGTTTGGCGAAGCTGGGCACCTCTTGAGAACATCTACACACACCTGTA	166425
Qy	1080	-----	1080
Db	166426	GTGGCTAGGCAGCCATGGGAGGGAAGAGCAGCAAGGAGATACTTCCACGGTGACT	166485
Qy	1080	-----	1080
Db	166486	AAGAGCAGCTGCTTAGAGGAGGGAGCGCAGCTCTTGATAGTTTCAGGAAGCCACCTA	166545
Qy	1080	-----	1080
Db	166546	GGTTTCTATTCTTGATGGTCTCACCACCACTAAAGTAATCCAGCCAGGGGCCACCGG	166605
Qy	1080	-----	1080
Db	166606	CAATGGCAATGTCAGCAGTGGTGGCTAGTAGCAGCTGACCTGAGGGACATGTGCAT	166665
Qy	1080	-----	1080
Db	166666	GTTTCTGAGCTTTGGAGAACAGAGAAAGGCAATTCCTGGCAGAGGACAGCGTGTGCAA	166725
Qy	1080	-----	1080
Db	166726	AGGGCCCAAAACACAGGAGTTGGGACACTTGGAGCCACAGGTCAAACTGTACACAGCA	166785
Qy	1080	-----	1080
Db	166786	AAATTACTACCCAGCAGCCAGGGAATCCCAAGTCACTGCTGCACTGAAATCCATGCCGA	166845
Qy	1080	-----	1080
Db	166846	GGTGACTCTCCCTAAGCGGGGGGGGGGGTATTGTAGSCTCACAGTTAGGGAGAGGT	166905
Qy	1080	-----	1080
Db	166906	TGGCCTACTGACTCTGGGAAGATCCCTTCGAGACCTGGGAGAGAACAGAACTAGGGA	166965
Qy	1080	-----	1080
Db	166966	GGGGGTAGAAAGTGGGGTGTGGCTTAGAGGCTGAGTGTGATCAAGAGGTGATGGTGA	167025
Qy	1080	-----	1080
Db	167026	GCGGTGAGATCCAGGACTGGGGTCTTGTGGAGGGCTCAGTCAATTAAGACAGAGTA	167085
Qy	1081	-----	1086
Db	167086	TATAGAGGAGCCCCCAGATATACACTTCCTGTTTCTTGTGAGGTACCTCTGCTCT	167145
Qy	1086	uGlyLeuHisAspThrAlaArgProValLeuProArgProThrIleSerAsnProPr	1106
Db	167146	GGGCTCCACGATAGTGGCCGGCGGTCTGCGACGTCCCCC---ATCTTAACCCCC	167202
Qy	1106	oProLeuIleSerSerAlaIleHisProSerValLeuGluArgGlnIleGlyAlaIleSe	1126
Db	167203	ACCCCTCATCTCTCTGCCAAGCACCCCTGGTGTACTTTGAGAGGCAGCTGGGTACCATCTC	167262
Qy	1126	r-----	1126
Db	167263	CCAGGTGAGTACCTCTTGGGCTGGGTCTCCCTTCACGGAGGCCTTCTGAGCCGTGGGA	167322
Qy	1126	-----	1126
Db	167323	TGTTGCCAGGGGTGGTGGTGGACTCAGCCCTGCTGCTGTGGGCTACGTGTGTACA	167382
Qy	1126	-----	1126
Db	167383	GGCCGTGTGTTCCAGTGGCTTTAGGCAGGTGACTCACTGTGTGAGCTTCAGTTTCATCT	167442
Qy	1126	-----	1126
Db	167443	CTGTTCACTGGGAGGTGGGAGCTGCAGCTGCTGAGGCTGTAGAGTGTCTATGCTCTGTC	167502
Qy	1126	-----	1126
Db	167503	GTAAGCCCTCGGTCACTCTGTCATCTTACTGACAGTCAACCACTCACCACATGGTTTGCA	167562
Qy	1126	-----	1126
Db	167563	AAGCAGTCACTGCTCTTCTGTAGACACTGGGCTTGGGCTGTGTTGGCTGGCCACA	167622
Qy	1126	-----	1126
Db	167623	GTCTCCAGTCAGGGCCGGCAGCCCGCTCTCTCACTTGGCTATCTTAAGTCCCTAAGCA	167682
Qy	1126	-----	1126
Db	167683	AGGCTCTTCTGTGTCTGAATCTGGTCTTCCGCTGGCTGAGTCTGGTTGTGATCTATCTC	167742
Qy	1127	-----GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAl	1143
Db	167743	TTCCGCCCTAGCAGGGAATGTCACTGCGCTGCGGCTGCCCTCAGAGCATGCCAGGC	167802
Qy	1143	aProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys	1160
Db	167803	CCCCATGGGCCCTCTCACTGAGGAGTCCCTCTGCTGTCGACCTTAAGAGCTGGGTGA	167862
Qy	1160	-----	1160
Db	167863	GTGCCCATGACCTTTTGGTTCTATGCCCCGGGGGCGAGTCTTGGCCCTGGGAACCTAGAG	167922
Qy	1161	-----LeuAlaProPheSerGlyV	1167
Db	167923	GGGGGGCAGCCTGAGACCTCTGTGTACCTCTCGCCCCACCCCTAGCCCCCTTCAGTGGAG	167982
Qy	1167	allysGlnGluInLeuSerProA:gglyGlnAlaaglyProProGluSerLeuGlyValP	1187
Db	167983	TGAAGCAGGAGCAGTTGTCCCTCGGGTCAAGGTAAGTCTGGCCACCTGAAAGTCTGGGAGTG	168042
Qy	1187	roThrAlaGlnGluAlaSerValLeuArg	1196
Db	168043	CTACTCTCAGGAGACCTCTGTCTAAGAGTAAGTCTGGCCCACTCACCTTCAGTCTTA	168102
Qy	1196	-----	1196
Db	168103	GAGAGCTGAAAGCCCACTTGACAGTGTGTCTCTTACCCTTGACTCCACAGGAGACCGCT	168162
Qy	1197	-----GlyThrAlaLeuG	1201
Db	168163	AGGTGGCTTTTAGACTTTGGTCTAAAGAGCTGTCTTATGCCCTTCTAGGGACAGCACTGG	168222
Qy	1201	lySerValProGlyGlySerIleThrLysGlyIleProSerThrArgVal	1217
Db	168223	GCTCCACTGCAGTGGAGAGCATCACCAAGGGCTCCCGAGTACCCGGAAGTGCAGAGTGCC	168282
Qy	1217	-----	1217
Db	168283	CCAGTACCAGGCTCTATCACCCAGTGTGTGCTGGGGGCGAGGTGGGATGGGCAGGCA	168342
Qy	1217	-----	1217
Db	168343	TCCTAGGAGGTGGGATGGGATGAGAGGACCTCCGGGGAGACGAGAGATCTTTGAGCTGA	168402
Qy	1217	-----	1217
Db	168403	ATGGTTGCTCACTGCGGTGTGGGTGAGGCTAGGCTAGGTGATGAAGAAAAACAGGTGTCTAT	168462
Qy	1217	-----	1217
Db	168463	GTCTCCAGCAGTTGGGTGTGGCAGAGACCAACCAAGAGGATAGCCAGGCCCTTAACC	168522

QY	1217	-----	1217
Db	168523	AGACTGTCAAGAGCCATAGATCATATAAGTGTCTGAGCTGAGAAAGTTACAGGTGACCCC	168582
QY	1217	-----	1217
Db	168583	AGTGACATGTTGCCATCGCCAGAGAAGACAGAGGGACAGAGGTGGCAGCTGGCTTTGA	168642
QY	1217	-----	1217
Db	168643	CAAAATCACCAGTCGGTAGGCATGAGCCAGTGGTGGGTACCAAGCTTAGTACCCAGGTGT	168702
QY	1217	-----	1217
Db	168703	GTGGCCAAAGCATACAGCCAGCGAACTTCCAGTCAAGGACCAACCCACACAGGGTCCTT	168762
QY	1217	-----	1217
Db	168763	TTGACTCACTGTAGGGACAAATGTCCATTATGCGCGAGGCTTCCCTCGAAAGTCTCTGG	168822
QY	1217	-----	1217
Db	168823	AAATCTTGGGCTTTCAAAGGCAAGCTCAGAACTGGAGGCGGAAATTCAGTTCATGTCTCT	168882
QY	1217	-----	1217
Db	168883	TCTGGCAGTCAGTAGCCCTCCTGACTCAGTGTTCCTGTGACAGAAATGTCTTTCC	168942
QY	1217	-----	1217
Db	168943	AGGATAGGTACTCAAGAGCTCTTAGAGTGCAGGTTTTCACACAGAGCTTCGCAAGGC	169002
QY	1217	-----	1217
Db	169003	TCAGGGATCGGGCTCACAGGCTGAGNATCCCTTTCCCTGTGTTAGGGCTTGTGAGAGT	169062
QY	1217	-----	1217
Db	169063	GAGGTCAGCGCTCTGGGGCTATCTGTGTGGATGATGACCCAGGCATTTGATGTGACAG	169122
QY	1217	-----	1217
Db	169123	GCTAACAGGTGGAGACCACATCCCACTGTTCCTGACGCCCTGGAACCCCAAGGCCAC	169182
QY	1217	-----	1217
Db	169183	AGAGCCCATGTCCCACTCAGCTCCCAAGNTTACATCAGGCGCTTGGAGCGGGCTGCGG	169242
QY	1217	-----	1217
Db	169243	TGGGCCCTCTGGCTCCCTGTATCTTCTTGGTTTGTCCAGTGAGTTATACGGAGCAGAAC	169302
QY	1217	-----	1217
Db	169303	CATGGCCTGCTGTATTTTGGTTGTGAGAGCCCTCTAAGCTTGGAGACTCAAAACATC	169362
QY	1217	-----	1217
Db	169363	AATACCCCTTTTCATCTGGAGCGTGCCTGACCTTTCTCACCGTATACCTGGCCCTAGCAA	169422
QY	1217	-----	1217
Db	169423	GCAGCTACAGCCACTTGAAGGCATTTCCCGACACACCCAGAAATCACCCCGCATGCC	169482
QY	1217	-----	1217
Db	169483	CCACGGGGACCCCAACAGAACTCCCAAGTCCCAAGGAGACAAATCACTGAATCTGTTCC	169542
QY	1217	-----	1217
Db	169543	TGAGGTCATAAATGAAGAGAGAACTCAGGATTTGGGGGCTCTGAGAGCCAAAGTTAGAGT	169602
QY	1217	-----	1217

Db 170683 AGGAAACACAGACTGTAAATGTGTCCCCAGACAGCAGTTGTGGTGTCTATACACTTTGC 170742
Qy 1281 ----- 1281
Db 170743 TGAGAGGGGAAGTCTTTTGGCAGAGCAGAAAGCAGCCAGGGCCCTTTGGGTTGTGTG 170802
Qy 1281 ----- 1281
Db 170803 TGGTAGCCTCAACAGGTTGAGGACTGGGCCCTAGGCCCTGCTGCCACCTTAGACATAGGG 170862
Qy 1281 ----- 1281
Db 170863 AAACGGGAGATGGACATAGTCACATTCACTGTCCTTCTATGAGGTAACCATCTGGCCC 170922
Qy 1281 ----- 1281
Db 170923 TGGTTAGACAAGCAACCTTTGTTAGGGAGGAGCCATGGGATCTATAAAGGTGTCTTT 170982
Qy 1281 ----- 1281
Db 170983 AGAGCAAGGCAGAGCTGAAGGACACACAGGAGAGCAAGGAGTGACTTCTGTCTCTGC 171042
Qy 1281 ----- 1281
Db 171043 AGGCATACAGCCCAAGGGGTGAGGGCTGGCCCTTGGTGGCTTCCCTACAGAGGTGGG 171102
Qy 1281 ----- 1281
Db 171103 AAGCTGGGGAATCTTTGAGTGGCGTGACAGGAAGGGCTGTGTGTAGAGCCATGACACCC 171162
Qy 1281 ----- 1281
Db 171163 TTCTCTGTGGCTAGAAAAGAGCCAGGGCTGCCTAAAGTTGAGGAACCCATACCTGGGA 171222
Qy 1281 ----- 1281
Db 171223 TGGCTCTCCAGTACCAAGCACCACATCAACTGCTGCCATGACTACTGCGAGGCTG 171282
Qy 1281 ----- 1281
Db 171283 GGGCTGCCAGTTCCTATTCTTTGAGTCTCAGAGCCCTCTGTGTCTAGAGTCAACAG 171342
Qy 1281 ----- 1281
Db 171343 TGGCCCATGGGCTTTGGCACCTGACATAAAGTACAGTCTTAAGGACAAGGGCTACAGC 171402
Qy 1281 ----- 1281
Db 171403 TGACCCCGGTGCTGGCTGCATGTCCAGGTGACAGACCCAAAGACAGCAAGGAGCGGTGAT 171462
Qy 1281 ----- 1281
Db 171463 CAGAGTCTGGCCAGAGCCCTGAGTCAGGAGTGGNGGCTCTAGAGTCTCAGGTTGTA 171522
Qy 1281 ----- 1281
Db 171523 ACTTTACAGCCTATCTTGGTGAGACTGGGTATCCTGTTCGTCTCTTTGGGGCCAGTGG 171582
Qy 1281 ----- 1281
Db 171583 GTCTGAAGCAGTTCAGAGTAGCGCCACCGTCAAGGCCAGGCCAGCTGCCAGCCCTTTC 171642
Qy 1281 ----- 1281
Db 171643 TTCGCAATCTCCATAGCCAGAGCATCAAGTCAATTGGGGATAGCAATGCCACTCTGTCC 171702
Qy 1281 ----- 1281
Db 171703 CCGACAGCTGAGGCTTCACATGGTCTTTGGACCGGGTGTCTCACACATCCCTTTGTATCA 171762
Qy 1281 ----- 1281
Db 171763 TGCAGACCTGCAGGGCTTTGTGGGGGCTGCTCTGCGGACTCAGACCGGTGTGACATCTTC 171822

Qy 1281 ----- 1281
Db 171823 GTTAGTGTGCATACATACCCATGGGCACAGACACATACACCCCTTACACAGCAGTGG 171882
Qy 1281 ----- 1281
Db 171883 CAGGCATGCTGGCACTTTCTGCAAGAGCAGTTCACAGTTGAGCTGTTTGGGGTCCC 171942
Qy 1281 ----- 1281
Db 171943 ATCTGTGGGGCAGGGCCCTCTTTCAAGTCTCCCTGTTTGGCTAACTCATCCCTGTGTT 172002
Qy 1282 ----- GlyMetSer 1285
Db 172003 GTCTGCACCTTGTGCCACCCACACCATGTGCTTTTCTGGCCCTCAGTGTGTGTCG 172062
Qy 1286 ValThrGlnCysSerIleGluAspGlyArgSerSerGlyProHisGluThrAla 1305
Db 172063 GTGTCACTGCTCCAAAGGAGTGGAGGAGCAGCTCAGGCCACCCACAGAGTGC 172122
Qy 1306 AlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAla 1325
Db 172123 GCGCTTAAACGACCTATGACATGATGAGGCGCGTGTGGCAGGACCATCACTCAGCC 172182
Qy 1326 SerIleGluGlyLeu----- 1330
Db 172183 AGCATCGAGGTAA--GTGTCTCACTGTAGTTCACGGGGTCAAGCACCATCCACGA 172241
Qy 1330 ----- 1330
Db 172242 GACCCAGGTACCCAGGTAAAGTAAGAAAGGAGGGGTGGCCCTTGACATGCTTGGAGG 172301
Qy 1330 ----- 1330
Db 172302 AGCCAGTCACTAGTGTGGTCAATATTAA CAAGTAGAGAAACTGAGGCACAAACCCGTGCA 172361
Qy 1330 ----- 1330
Db 172362 AATGAGCCCGCAGGAAACGGGGCACTTGTCTGCTGGTGTAAAGTCAACAAGAGACTG 172421
Qy 1330 ----- 1330
Db 172422 TTGGTGTAGAGCCCGCCCTCTGGACACCATACGTTGACACTATCCCTGTCTCCCCA 172481
Qy 1331 ----- MetGlyArgAlaIleProProGluArgHisSerProHisIleLeuLysGluGln 1348
Db 172482 GGACTCATGGGCGCTGCCATC---CCTGAGCAGCAGAGCC---CACCTCAAGGAGCAA 172535
Qy 1349 HisHisIleArgGlySerIleThrGln----- 1357
Db 172536 CATCATCCGAGGCTCCATCAGCAAGGTATATCCTCTTACGGTGTGTAAATTCACCT 172595
Qy 1357 ----- 1357
Db 172596 CCAGCCAAACACAGCTAAGGGGTGGAAACCGAGGGTGAACCTTCCGTCCTGTCTTAT 172655
Qy 1357 ----- 1357
Db 172656 AGTGGAGTCACTACTCTGTACTTGGGCTCATTTCCAAAGCCCATGTCGATGCCCTTGT 172715
Qy 1357 ----- 1357
Db 172716 AGTGATTGAGAGGGGTCCAGATTGAATGTCTTTAGCTCATGCTTTGTGTAAGTCTG 172775
Qy 1357 ----- 1357
Db 172776 AGCAATGTGTTTCACTTACGTGCTCCTCAGCTCATATGAAGAAACAGGCTTAGAAAAAGTCA 172835
Qy 1357 ----- 1357
Db 172836 GTTTGCATACCAAGGTCCACAGGAAGGATTTGTGTGACCCAGGCTGAGCCAGGCAAG 172895

Qy	1358	-----GlyIleProArgSerTyrValGluAlaGl	1367	Db	173975	TCTGTACCCCTCCACAGTGGGGTCTCTGGAGTATAGCCAGGCTGAAGCAGCAGGG	174034
Db	172896	GCACTGGAGACCAATGATGTTGCCCGCAGGACATCCCGAGGTCAATGTGGAGGCCA	172955	Qy	1452	-----	1452
Qy	1367	nGluAspTyrLeuArgGluAlaIysLeuLeuIysArgGluGlyThrProProPr	1387	Db	174035	TAITAGCCTCTGACGTCTGTTGCTCAAGGGTTAGAGATGCCCATCTCTGTGCAGAC	174094
Db	172956	GGAGGACTACTTACGTAGGAGGCCAAGCTCTTGAAGCGAAGGACCGCACCGCCCC	173015	Qy	1453	-----GlnGlyTh	1455
Qy	1387	oProProSerArgAspLeuThrGluAlaTyrLysThrGln-----AlaLeuGlyPr	1404	Db	174095	ATACAGGAGAACAGCATCTGGCCAGCCACATACAGACCTGGACCTCTCCCCAGGGAC	174154
Db	173016	ACCGCCACCTCGGACCTGACTGAGACTACAAACCGGCCCTCGACCTCTGGGTCC	173075	Qy	1455	rProLeuIysTyrAspThrGlyAlaSerThrThrGlySerIysLysHisAspValArgSe	1475
Qy	1404	oLeuIysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaArgSe	1424	Db	174155	CCCACTCAAGTACGACTCCCGGGTGCCTCAGTGGCTTCCAAGAAACACGACGTACGCTC	174214
Db	173076	CCTGAAGCTGAAGCAACTCATGAGGTGTGTAGCGACTGTGAAGGAGCGGCCGCTC	173135	Qy	1475	rLeuIleGlySerProGlyArgThrPheProValHisProLeuAspValMetAlaAs	1495
Qy	1424	rIleHisGluIleProArgGluLeuArgHisThrProGluLeuProLeuAlaProAr	1444	Db	174215	CATCATCGGAGCCCGCGCGCTTTTCTGCGCTGCACCCACTGGACATAATATGGCTGA	174274
Db	173136	TATCCATGAGATCCACGAGAGGAGCTGCGCCGACCGCGAGCTGCCCTTGGCACCACG	173195	Qy	1495	pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuIysSerArgProGlyThrAl	1515
Qy	1444	qProLeuLysGluGlySerIleThr-----AlaLeuGlyPr	1452	Db	174275	TGCCCCGGCACTAGAGCGCGCTGCTATGAAGAGAGTCTGAAGAGCCGGTCAGGACCA	174334
Db	173196	GCCTCTGAAGAGGAGTTCCATCAC-CCAGGTACGAGGACACAAAGAGGGCGGGGCT	173254	Qy	1515	aSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyL	1535
Qy	1452	-----	1452	Db	174335	CAGTGTGTCAGGGGGTCCATCACACGCGGGGCCCCAGTTGTGCTGCTGAGCTAGGCAA	174394
Db	173255	TGGCAAGCCTGAGGGAGTCTAATGCTGCCCTCATTTGCTTGGTTTATAAACTAAGGCC	173314	Qy	1535	sProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr	1555
Qy	1452	-----	1452	Db	174395	GCCCCGCAAAAGCCACTGCGCTACGAGGACCAACCGGGGCCCTTTCACAGCCACCTGCC	174454
Db	173315	CAGAGAGTCAGCGGCTTGCTCTGCTCACTGACAGGCTTTGTGAACCAAAAGTCTTTGG	173374	Qy	1555	oArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSe	1575
Qy	1452	-----	1452	Db	174455	CCGAGGCTACCTGTGACTACAGGAGGCCACGCCACGCTTCAGGAAGGTGAGGC-TC	174513
Db	173375	TCACCTGGCTGCTCCCCACATTGCTACTGTCATGTTGTGTGTCAGACTAGAGCCACTGT	173434	Qy	1575	r-----	1575
Qy	1452	-----	1452	Db	174514	CCCTGGGCTGCCAGCTGTGGGAGGGGACAGCTCATTTTCCACGCGCAGTCTGTCTGCTG	174573
Db	173435	CCCCTACCTTGGTCATGGCTTGATGGAAAGAAAGACAAACAGCCAGACCACTA	173494	Qy	1575	-----	1575
Qy	1452	-----	1452	Db	174574	CCCTGCTCCACACTGTCATACAGAGCCCCCGTGTCCCGCTTACCAAAATATTTTAA	174633
Db	173495	GGGAAGCTGTGTCTACCTGAGCCAGGACACACAGTGATGGACCAAGATGATTTAGG	173554	Qy	1575	-----	1575
Qy	1452	-----	1452	Db	174634	TAATTCAATCAGGGCTGAGGAGGGGCTCAGTAGCTAAAGTCTTGTGTACCAAGCATGA	174693
Db	173555	GTATCTTTTGAATCTTAGTGTCTCTCTCATTTGGGCCAGCACTTGGAGGGCGGAG	173614	Qy	1575	-----	1575
Qy	1452	-----	1452	Db	174694	GACTCAGAGTTCAAAATCCCGGTATCCACATAAGCCAGGACACTCATACATCTCTGTA	174753
Db	173615	TCAGGAAGATTGTAAATTTGAGGCTAGCCTAGGCTACAAAGCTGTACTCTGGCTTACTGC	173674	Qy	1575	-----	1575
Qy	1452	-----	1452	Db	174754	TTCTCATGTTCACGTGAGACAAAGGCTAAGACAGAAATCTCTCTAAAGGTTATAGCCA	174813
Db	173675	CGATCACCAAGTTTACAATTTAATGACAAACAAACATTTGAGTGAATTTGAGATCTGG	173734	Qy	1575	-----	1575
Qy	1452	-----	1452	Db	174814	TGAGTCGGGTCTCAAACATGGCGGAAAGTGGATGACACCTCTCCCCCCCCCTCTATT	174873
Db	173735	GACCTAGCTGTCTACTGTCCCGAGCCTGGATCCCTGTGGGATCTCTCCAAGATGTC	173794	Qy	1575	-----	1575
Qy	1452	-----	1452	Db	174874	TTATACACAGCATAGCACACTCATATCAGCACACTCGCGCCTACATTTCTGTGTGCAC	174933
Db	173795	AATAATTTCCCGTCTCCGTGGTGTCTCTGAGAGCCACAGGAATGTGTAATCCAGCCT	173854	Qy	1575	-----	1575
Qy	1452	-----	1452	Db	174934	ACATGTACACAAAAACAATAGTTTAAATGAAGCAGTTCCCACTCTGTAGGCTTACCT	174993
Db	173855	GGGACCGGAGCCAGAGCTGGGGTGACAGCAGATGCATCTGATATGAGTCCATGAGTCT	173914	Qy	1575	-----	1575
Qy	1452	-----	1452	Db	174994	CTCCAGTAGCGTGTGGTCTGAAGACACCTGCTGTGACGTGACCGCGGCATCTCTACTCT	175053
Db	173915	GCAGGATGTCCGGGGCCACACTGGGTGAGGAAGGTGCAATTTGACCTGTAGTGTAGCCACG	173974	Qy	1575	-----	1575
Qy	1452	-----	1452				

```
Db 175054 GAATCCGCTTCTGATTCTCTCGCATTTCTGACCTCCGCTCTTCTCCGATCCTCCACAT 175113
QY 1575 ----- 1575
Db 175114 GTGAGCCTTTCCGTGTGCTGGCCCTCTATGCGAAGGCTCCACCTTTGGTTTGTCTTCT 175173
QY 1575 ----- 1575
Db 175174 AGCCAACCTGTATACCCACGTCACCATGTCAGGGCTCGAGAGGCTATAGGCACC 175233
QY 1575 ----- 1575
Db 175234 TCCAACCTAGTTCCAGCCACATCCGGGCCGGATTGTGTAGGGGTGCAGACCAGTGTGTTT 175293
QY 1575 ----- 1575
Db 175294 AATTCAATTGATGAGGAGGCCCTTCCAGAGCACTGAGATCCTCAAAACATCTCTCTCTCT 175353
QY 1575 ----- 1575
Db 175354 CAGGCCGTGCATCTCTACCTTGGGTTTCATCTCTATCTGGAACAATGGCTCCTTGGAG 175413
QY 1575 ----- 1575
Db 175414 CTCAGGGTCTTGCTCTCTCTCTCAGGGTGCTCCCTCTTCTCTCTCATGGAACTGCC 175473
QY 1575 ----- 1575
Db 175474 AGCCCATGATTTAAGTCAGACTGGAATGAATAGGGCAATGGTGACTGGGGAGTAAATGAA 175533
QY 1575 ----- 1575
Db 175534 TGCTTTCAAGGAGCCAGGAGCAGAGAGAGAGAGCTCTAGGGCTTTGTGACAGCTCAG 175593
QY 1576 ----- 1576
Db 175594 CCTGTGCGCCGTTGAGGCTGAGGCCACTTTGACCTGTGCCCCACAGGCCTCCCTCTC 175653
QY 1576 rSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPr 1596
Db 175654 CAGCAAGCGCTCCAGAGCCGGAAGCTGACCTGCACACCCCGGAGATTGCCAAGTCCCC 175713
QY 1596 oHisSerThrValProGluHisProHisProIleSerProIleSerProIleHisLeuLeuAr 1616
Db 175714 ACATAGCACTGTGCCCCAGACACACCCCCCACTCCCTCCCTACGAGCACTGTCTCCG 175773
QY 1616 gGlyValSerGlyValAspLeuTyArGserHisIleProLeuAlaPheAspProThrSe 1636
Db 175774 GGGTGTGACTGGTGTGACCTGTACCGTGGCCACATCCCATTTGGCCCTTTGACCCCACTC 175833
QY 1636 rIleProArgGlyIleProLeuAsp----- 1644
Db 175834 CATACCCCGAGGATCCCTCTGGAAGCAGGTGGGTGCTGGGCTGTATCATGTGGGGCTT 175893
QY 1644 ----- 1644
Db 175894 ATTGATTGTCTCTTGGGTGGACACTTGGGGTATGGCTTAGTGTGCCATCACCAGCT 175953
QY 1644 ----- 1644
Db 175954 GCTTACCCAGGCAGGTCTTAGGATTTGGGCTTTTGTGCGTGTCTTCACCGTAGGACT 176013
QY 1644 ----- 1644
Db 176014 CATCTGCTCCCTTGCAAGTCAGTAAATGAGGGGCTGCAGTGCAGTGCATGATGGAACC 176073
QY 1644 ----- 1644
Db 176074 TTTTAAAGGCTCCTTTGTGAGGTGTAGCACCTGTATTTTCGATGCATTAATTGGGGG 176133
QY 1644 ----- 1644
Db 176134 GCACCTCCGAAACATTAATAAGCAGGCCCCACATGATGATGAGGAGCGAGTCTTAAGCTC 176193
QY 1644 ----- 1644
QY 1644 ----- 1644
Db 176194 TGTGTGTGCAATTGGGAGGCATGTGGTGTACAAGGAGAGAGCTGTTGTGSCAGAGT 176253
QY 1645 ----- 1645
Db 176254 GAGGAACGTGAAGATAAGTCAATGAGGCCACCTGACCCCTCCATCCTCCCTCCCTCCACG 176313
QY 1645 ----- 1645
Db 176314 AGCCGAGCCTACTACTCTCCCGGCACTTGGCCCCCAGCCACCTACCCACACTGTGA 176373
QY 1665 rProProTyLeuIleArgGlyTyProAspThrAlaAlaLeuGluAsnArgGlnThrIl 1685
Db 176374 CCGCCTTACCTCATCCGGGGTACCTGCACCCGGGGCTCGAGAACCGCCAGACCAT 176433
QY 1685 eIleAsnAspTyIleThrSerGlnGlnMetHisAsnThrAlaAlaMetAlaGl 1705
Db 176434 CATCAATGACTACATCACCTCGCAGCAGATGCACCAACAGCAGCCTCTGCCATGGCCCA 176493
QY 1705 nArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerSerLeuAlaLeuAsnTyAl 1725
Db 176494 GCGCGCGGATATGCTGAGGGGTCTGTCCCGAGAGAGTCTCGCTGGCCCTCAATTATGC 176553
QY 1725 aAlaGlyProArg----- 1729
Db 176554 CGCTGGCCCGCTGGTGTGAGTTGCTGGGGCTGTGCCCATCAGCAGACTTGTCTCTGGC 176613
QY 1729 ----- 1729
Db 176614 TGTGGTGTGCAAAAGAAGAGTGACACGTGTGCACCTGCGAGTGGCGGAGGGGCACAGTT 176673
QY 1729 ----- 1729
Db 176674 GGCTGGGAGGACTAGTGTGTGATGCTGCTCTTGGGGAAGGCAGCAGCGGGCCACTT 176733
QY 1729 ----- 1729
Db 176734 GTATCTGTTCACTATAGAACTGTGTGTAGCAGACACTCTTGCCACCATGCACCGTGGGT 176793
QY 1729 ----- 1729
Db 176794 CTTTGGTGTGTTCTGTAAACCTTGGTATGGGCACTCTTCTGTCTATCCAAGAGATCAAAA 176853
QY 1729 ----- 1729
Db 176854 TACAGGCTGGATGTGGCAATTGCACGCTATATCTTAAGTACGAGGAAGGTGCAAGGCCAG 176913
QY 1729 ----- 1729
Db 176914 CCTCAGCTACATAAATTTTGGAGCCAGCCTGGGCTATATAGTCTGCCTCTATACAAATG 176973
QY 1729 ----- 1729
Db 176974 ACTGAAAAGAAAGAAATGAGAGCATGCGGCTTGTCTAGAACACTGACTTCTCTGTGAACAGC 177033
QY 1729 ----- 1729
Db 177034 CTGGGTCACTTTGCTCGACTGTTTCAGAGATAACCTTGTGGCCCCAGTTAGTGATACGGA 177093
QY 1729 ----- 1729
Db 177094 AGCCACTCTGGGGTTTTAGAGATGGGAGCCCAATGCAAGGAGCTGATACTGTGGGAAG 177153
QY 1729 ----- 1729
Db 177154 GCAGTGTTTGGGTCAAAAGATGTGTGTTTGAGAGACACTGCACGTCTATAGCCTTGAGAAA 177213
QY 1729 ----- 1729
Db 177214 GGGGCTTGGGGCCACTCTCTGGCACCCAGGAGGGGTAGTGAGGTGAAGGTCTCAGGCCT 177273
```


Qy	1730	-----G	1730
Db	177274	GCACCTTTGGGACAGGACATCAGGCACCTACAGCTGACATCCACCCACCCCTCGAC	177333
Qy	1730	lyllelleAapLeuSerGlnValProHisLeuProValleuValProThrProGlyT	1750
Db	177334	GCATTATCGACCTGTCTCCAGTGCACACCTGCGGTGCTGGTCCGCCCACTCCAGGCA	177393
Qy	1750	hrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerS	1770
Db	177394	CCCTGCGCACCGCATCGACCGCTTGCCTACCTCCCACTGCGCCGCCCTTCAGCA	177453
Qy	1770	erArgHisSerSerProLeuSer-----	1778
Db	177454	GCCCTCAGCAGTTCCACCACTGTCCCAGGTGCGCTGCCAATGCTGGTGGGTCT	177513
Qy	1778	-----	1778
Db	177514	GTGCTCACCCCTGTGCCCTGCAAGTAAGTTCTGGGCCACTGGGCTCTGGCTCACACTC	177573
Qy	1779	-----ProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerS	1794
Db	177574	TGGCTTTGGGGTTTCCAGGAGGCCCACTACCTAGCTTAACCACTGCCACGTCATCAT	177633
Qy	1794	erGluArgGluArgAspArgGluArgAspArgAspArgGluArgGluLysSerI	1814
Db	177634	CGGAGCGGGAACGGGAACGAGAGCGAGAACGAGAC-----AAGTCCA	177675
Qy	1814	leLeuThrSerThrThrThrValGluHisAlaProLleTtpArgProGly-----	1830
Db	177676	TCCTCAGTCCACCATACAGTGGAGCATGCCACCATCTGGAGACCTGGTAGGSCACCCC	177735
Qy	1831	-----ThrG	1832
Db	177736	AGACCCCGCCACCCCGAGTCTCATGGGCCACCTGGGGCTGCCCTGACTCGGCGG	177795
Qy	1832	luGlnSerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProA	1852
Db	177796	CCTG-TCAACGAGTACGAGCAGAGCAGCGGGCTGGGGG-----AGCAGCGCCCG	177848
Qy	1852	laSerHisSerHisAlaHisGlnHisSerProLleSerProArgThrGlnAspAlaLeuG	1872
Db	177849	CTTCCCAACAC-----CACAGCACTCGCCCATCTCCCGCGGACCCAGAGCGCTTGC	177902
Qy	1872	lnGlnArgProSerValleuHisAsnThrGlyMetLysGlyllelleThrAlaValGluP	1892
Db	177903	AGCAGAGCCCGAGTGTCTGCACAAACAGAGCATGAAGGCGGTGTCACCTCCGTGGAAC	177962
Qy	1892	roSerLysProThrValLeu-----	1898
Db	177963	CCGSCACGCCACCGTCTCT-GAGGTGGGCCAGGTTGGCACCCAGGGGGAACAAGCCTGGGG	178021
Qy	1898	-----	1898
Db	178022	GGAGGGTACACAGATGGATTTGCTGAGAGGAGGGGTGTGGGGGTGGGCAGATGGGTAGG	178081
Qy	1898	-----	1898
Db	178082	GAATGGAGCAGATGGCAGGATCGGAAGGTGGGAAAGTGAGAGAGACTATGAGAGTGGGA	178141
Qy	1898	-----	1898
Db	178142	TGGGTGCTTGTACAGCGGGGAGGATGGCCAGGGGAGAGGGCACACAGTGAGTGGACG	178201
Qy	1898	-----	1898
Db	178202	GGAGATAGATCAGATAACAGGAGATGACAGTGGGTGGATAAAGGATGGATGGATGGA	178261
Qy	1898	-----	1898
Db	178262	TGGATGGATGGATGGATGGGTGGGTGGGTGGATGGATGGATGGATGGATGGGCGAGA	178321
Qy	1898	-----	1898

Db	178322	CAGATGGACGAGTGGATGGATGGACTAAATGGATGGATGGACAGACAGACAGGAGGAGA	178381
Qy	1898	-----	1898
Db	178382	CAGACAGATGAACGAGTGGATGGATGGACTGGCGGATGGATGGACCGACCGACGGATGGA	178441
Qy	1898	-----	1898
Db	178442	TGGATGGACAGGCGACAGACAGGCGACAGACAGATGAACGAGTGGATGGATGGACTGG	178501
Qy	1898	-----	1898
Db	178502	CGGATGGATGGACCGGACCGGATGGATGGACAGGCGACAGATGAGCGAGTGGATGGA	178561
Qy	1898	-----	1898
Db	178562	TGGACTGATGGATGAATAGACAGACCGGACGCATATATGAGTAGACAGCAAGTTGATGGGC	178621
Qy	1898	-----	1898
Db	178622	GSCGGATGAATGGGTGGATGGATGGATGGATGGATGGAACAGCTGGATAGACAGGT	178681
Qy	1898	-----	1898
Db	178682	AGACAGACACAGATGAATGNN	178741
Qy	1898	-----	1898
Db	178742	NN	178801
Qy	1898	-----	1898
Db	178802	NN	178861
Qy	1898	-----	1898
Db	178862	NN	178921
Qy	1898	-----	1898
Db	178922	NN	178981
Qy	1898	-----	1898
Db	178982	NN	179041
Qy	1898	-----	1898
Db	179042	NN	179101
Qy	1898	-----	1898
Db	179102	NN	179161
Qy	1898	-----	1898
Db	179162	NN	179221
Qy	1898	-----	1898
Db	179222	NN	179281
Qy	1898	-----	1898
Db	179282	NN	179341
Qy	1898	-----	1898
Db	179342	NN	179401
Qy	1898	-----	1898

Qy	2067	spLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuA	2087	Db	182689	CAGGGCTTACAACATGGCTCCAAGCACCCAGCAGCACTGTGTGGGCATACCCCTCTGG	182748	
Db	181610	ACAAGGGGCTCTCAAGCCTCTGGAAGAGCTAGAGAAGACCACTTGGAGGGGAGCTGC	181669	Qy	2204	-----	2204	
Qy	2087	rgProLysGlnProGly-	2092	Db	182749	ACAGGCTGAGGCAAGAGCACTGTGAACCTGCAGGGCAGCCCTGGGCTGTACCACAAGACCT	182808	
Db	181670	GGCACAGCAGCCAGGTGACACCCGGGAGAGTAGCAGGTGGGAAACTGAGGGCAGGCTC	181729	Qy	2204	-----	2204	
Qy	2093	-----	ProValLysLeuG	2097	Db	182809	GACTCAAAATAAAACCAAGCCCATGAGTGTGGGTCTCGAAGAGGTCTCAGAAGGCC	182868
Db	181730	CATGTTGCTCCTCAGGCCCTTGACAAGCTCTGCTGTGTCTCAGCCCCCATG-AGCTCA	181788	Qy	2204	-----	2204	
Qy	2097	lyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSers	2117	Db	182869	CCTGGGCTCTCCATACCTCAGATGCAGAGTCTGGCCATTACAGGGAAACACGTCAGGCC	182928	
Db	181789	GTGGGAGGCTGCCCATCTCCCATCTGGGGCCTCTGCCGAGAGCCAGCCCTCATCCA	181848	Qy	2204	-----	2204	
Qy	2117	erProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaG	2137	Db	182929	CACCTGGGGCAGTAGGGGTCTCTTCATGAGTAGAACCCCACTCTCAGCTGGGATGCT	182988	
Db	181849	GCCCACTCTCCAGACTGCCCCAGGCATCAAGGTCAACAGAGGGTGGTCACCTTGGCTC	181908	Qy	2205	-----	2205	
Qy	2137	InHisLysSer-----	2140	Db	182989	GTCTCCATTAGGTCTCCAGAACCCAGCAAAACA--GTCCTGAGTAGCAGCGAGGTGCT	183045	
Db	181909	AGCACATCAGTGAAGTGCCCATCTCTGGCCCTGCAATTGGGAAGCAGGTGGCTATGA	181968	Qy	2222	IleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVal	2241	
Qy	2140	-----	2140	Db	183046	ATTGAACCTGTGTCCCAACCGGAGGGCATGACTGAGCCAGGACATGCACGGAGCATCGG	183105	
Db	181969	CAGCCATATCCACAGCCAGACAGATGCTAAACATGCTTAACATGGCCACACCGTTAGTCC	182028	Qy	2242	TyrProLeuLeuTyrArgAspGlyGluGlnThrGlu-ProSer-----	2255	
Qy	2140	-----	2140	Db	183106	TACCCACTGTGTACCGGACGGGAGCGGGAGCCAGGTAGTATGGCACAAATTC	183165	
Db	182029	TTTGGGGTTCATATAAGGTCTAACCGAGGACACTTACCAGGAGACCAAGAGCCAGATC	182088	Qy	2255	-----	2255	
Qy	2141	-----	GluValIle	2143	Db	183166	CCGTTAGAGGTGGTGGTGTTCCTCCCATACACCGTGGCCTCAGCCCTCTAGCGTGG	183225
Db	182089	AGGGTTCTGCCCTACCTACTCTGAAAGTCTCTGAAGTCTGCTCCCTCTCCCTGAGAGGTCAAT	182148	Qy	2255	-----	2255	
Qy	2144	ThrGlnAspTyrThrArgHisHisProGlnLeuSerAlaProLeuProAlaProLeu	2163	Db	183226	AGGTGCTCAATCTCTTACAGCTCAGCGGCTGATGACTTTTCAGGGGGCTTCACTGGTGG	183285	
Db	182149	ACACAGGACTACACAGCCACACACCCACAGAGCTCAGTGGCCCTCTCCCGCCCTCTC	182208	Qy	2255	-----	2255	
Qy	2164	TyrSerPheProGlyAlaSerCysProValLeuAspLeuArgProProSerAspLeu	2183	Db	183286	CACCTGTGCTGGCATCTCAGCCCGAGTCAATCTTATTTATCATCTTGTGCTGTGATGA	183345	
Db	182209	TACTCTCTCCCTGGAGCCAGTGCCTCTGCTGATCTCCGCGTCCACCCAGTGACCTC	182268	Qy	2255	-----	2255	
Qy	2184	TyrLeuProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGly	2203	Db	183346	GAATGAACCTAAACGGGACACAGAGTGGGTAGGGTGGGCATGGACCTCAGGGAGGGTATGA	183405	
Db	182269	TACTCCACCCCGGACCATGGCACCCAGCCCGGGGATCCCCCAGTGAAGGGGC	182328	Qy	2255	-----	2255	
Qy	2204	-----	2204	Db	183406	CCTATGGCGGCCAGTCCCTGGGAATTTCTTGTCTCAGTATAGCTCTAAGAAGTTCTCAG	183465	
Db	182329	AAAAGGTGAGGAAGTACCCTATGCTGTCATCAGGTGCTGGGAACAGTGGGACAGGGGTGAG	182388	Qy	2255	-----	2255	
Qy	2204	-----	2204	Db	183466	GGGGAATTCACAGTGTGCTGGGAGGCTCAAAACCCGCCCTAAGCCAGGGCTAAGTACAG	183525	
Db	182389	CTTTGACGTGTGGGCTGATGGGAACCTGAGGTGCAGGGATACAAAGCCTCACTACGCTC	182448	Qy	2255	-----	2255	
Qy	2204	-----	2204	Db	183526	ATCCACAGCGCTGGGATGGGAAATGGTTGGTCTTTAGCACCTCTCAGTGTGGCTCAAGTG	183585	
Db	182449	CCAGAACAGACACAAAGCATCTAATCAAGAAGTGGGTCAACCCAGGGTCCCAGGTGCCA	182508	Qy	2255	-----	2255	
Qy	2204	-----	2204	Db	183586	GTGAGCTAGCCCTGTTGCTAAACACAGTGGCCAGCCTGCTGACCCCTGCGCTCTCATGC	183645	
Db	182509	ATTGAATGCTCTATATATACCACTCAGCTAGGGGTGAGATGTCGCCATCTTCCATTTAGCA	182568	Qy	2256	-----	2273	
Qy	2204	-----	2204	Db	183646	TCCTGACAGTGGGTCTAAGTCTCCAGGCAACACAGCCAGCCGCGCAGCTTCTTCAG	183705	
Db	182569	GATAGGAAACTGAGGCATAGAGAGGTGAGTTACTATCCGTCCCATGTCACACAAACACC	182628	Qy	2273	rLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluLeuAsnLysLy	2293	
Qy	2204	-----	2204	Db	183706	TAAGTGCACGAGAGCACTCCGCCATGTTGGTGAAGTCAAGAGCAGGAGATCAACAGAA	183765	
Db	182629	CTGCACAGATGTAGACAGGTATCTCTGGGTAAACAGCAGAGGTCTTAGCATGGGGGCTCAT	182689	Qy	2293	sLeuAsnThrHisAsnArgAsnGluProGluTyr-----	2304	
Qy	2204	-----	2204					

183766	Db	ACTCAACACCCCATTAACCGGAATGAGCCCTGAATACAGTAAGGAGAACCCAGCCAGAGCA	183825
2304	Qy	-----	2304
183826	Db	GGCACTGGGAGACAGAGTGCTCTTAAAGCCTTCCTGTCGAAAGGCTTGATAGTCTGCCAGT	183885
2304	Qy	-----	2304
183886	Db	GGATACCGACGGCATATAGCACCTCAGTAATGCCCTACGCACAGTAGGTGCTCAGCGAAA	183945
2304	Qy	-----	2304
183946	Db	CCTATAGAGACAGCCCTGTGCTGGCGCGAGGAGCACAGTGCACTACTCAGGTTTCATGC	184005
2304	Qy	-----	2304
184006	Db	TGCAGGCATGGCGCTTCTCCACAGCTGATGTGGTGGACAAGAACCCCTTGTCCTTAGC	184065
2304	Qy	-----	2304
184066	Db	TCAGGGTCGAGGGACCATATGCTTATGAGAAATGCACTGGGAAACAACACTTGAGGGCTT	184125
2304	Qy	-----	2304
184126	Db	CGCACTGTCCCTCCCAACAGCAGCCCTCCCTAGGGTGGCTTGGACAAATCCTTCTTCGG	184185
2304	Qy	-----	2304
184186	Db	AGTTAGACTACAGAAGCCCTTGACGGTTAGCAAGCCCTAGGCTAGTGACCTACTGCGCAG	184245
2304	Qy	-----	2304
184246	Db	AGTAAACCTGGAGCATCAGATACCTAGTTGTGCCAGACATTCAACCCAGACGCTACAGG	184305
2304	Qy	-----	2304
184306	Db	AGTAAACGGGTGCAGATCTCCAGAACCTGTGAGCTCAGGCCTTACCACCTTGTATACC	184365
2305	Qy	-----AsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyTh	2322
184366	Db	GGCTAGATATCGCCACGCTTGGACGGAAATCTTCAACATGCCCGCCATCCTCTGGAC	184425
2322	Qy	-----	2322
184426	Db	AGGTAAACCTCCCGTGTGTCGATACATCTCCCTTCCCAACTACTGACTTAGTGA	184484
2322	Qy	-----	2322
184485	Db	CAAGGAGCCTTAGCCAGTTCAAAATTGAGCCCTCTCCCTGGCCTTAGTGGACCCAGAATG	184544
2322	Qy	-----	2322
184545	Db	GCACAGACGCCAGCCTCCACTGCCTCAGTCTCCCTACATGAGGGAGCTCCAAGTTT	184604
2322	Qy	-----	2322
184605	Db	GAGATGAGGGTTTCTCCGGTGGGTATTAGGTACACCCTGGTAATACTTGCCCAATTGCT	184664
2322	Qy	-----	2322
184665	Db	GGGGCTGAGGACGTCACCCAGTGTGGGCTAGGAAATTGGCCTTGGGTCAGAGCTTC	184724
2322	Qy	-----	2322
184725	Db	AGGTGCCACAGGCATCTCTCACACTCAAACTCCATTTGGGTTTCCAGCAGAGAGCCCT	184784
2322	Qy	-----	2322
184785	Db	GGCCTTTTCTAAATCAGATGCAACGGGTAAATGTGAGAGAAACATGTCACTCAGGTG	184844
2322	Qy	-----	2322
184845	Db	CAGTTGTGGCAGAAATTTTCAGGCATTCAGTGTGTGTGTGGACTTCAGAAACATCGGCTT	184904

QY	2322	-----	2322
Db	184905	GCCATGGTTATGATGACGTTGCCCTCGGAAGAAGTGTTCATGGGAATAAATGCCTCAG	184964
QY	2322	-----	2322
Db	184965	GAACACCTGACCGGGA CTGTGA CACTGGGGCAGCTGTGAGGACTTGCACTCCAGGCTGG	185024
QY	2322	-----	2322
Db	185025	AGGAACAGGCAAGGTTCACTACCTTCCTTATGTGTGAGCTTTGGGCCAGGGAATAAG	185084
QY	2322	-----	2322
Db	185085	GTGCCAGATACACCCTTGGGAACAATTCCAGAGACCACAGGACCTCGAGGGGGTGG	185144
QY	2322	-----	2322
Db	185145	GGTGGGATCTGGCAGGGGCTGTCAAAACCCAGAAACATACTGGGGACTTTTCTGAGC	185204
QY	2322	-----	2322
Db	185205	CTCGGCAGAGGAAA CTTTGACAAGTTGAAGATGAATGTTTCTAGAAATGTTCTAAATCT	185264
QY	2322	-----	2322
Db	185265	TGATGAGGTTCTAGACACCTGGTGTGTGTGAGTAGAGATTACAGGGTAATACACGC	185324
QY	2322	-----	2322
Db	185325	ACACACACACACACAGTGTGCCTCCTCAAACACCCCCATAAATGTTAGTCTCTCAAAGT	185384
QY	2322	-----	2322
Db	185385	CATGTTTCCAGGCATGGCCAATGACACTCAACATCGGGCCAGAGTAATAATGAGGAAGTCT	185444
QY	2322	-----	2322
Db	185445	ATGCTGTGCCAGTGAGCGCCTCAGGTACAAAGGCCAGGAGCGGTCA GTGCTCCATC	185504
QY	2322	-----	2322
Db	185505	AGAAACTCCCCTCCACCCCTCTTTCCCTGTTA C TCTCGTATCTCCTCTGTATCTTG	185564
QY	2322	-----	2322
Db	185565	GATGTGGAGAGTGGTGGGCTTGTTTATTGAGTTAGGGCGTGGAGGGTGTGTGTG	185624
QY	2322	-----	2322
Db	185625	TGTGCGTGTGCGTGTGCGTGTGGCATGCAATTTACCCCTGT CAGAAAGTTTCGGCTTTT	185684
QY	2322	-----	2322
Db	185685	TC TTGGTAACATTTGCTTTAGCACTGAGGGAACGTGAGCATCCCTGACCCGAGCTTGCCCG	185744
QY	2322	-----	2322
Db	185745	AAACCCAAACTGAGGCTCCACAGCCACCCAGAACAGGATTAGGAAGTGACTGTCCCTTC	185804
QY	2322	-----	2322
Db	185805	CATCCTCAGAACCTGGCTATTTTACATFGCCGCATTTACGCCATGTGATTTTCACAAG	185864
QY	2322	-----	2322
Db	185865	GTTTGGTAGCCTGGAGGTTACGGGGCACAGAAACTGGCCAGTGGCATAGCTGGGAAT	185924
QY	2322	-----	2322
Db	185925	CCTCCGATGACGTGTTGTGTGTTTTTTTGTGTGGTGGTGAFTTCGGGGTTTTTTATTG	185984

TITLE
JOURNAL

Direct Submission
Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Mar 27, 2003 this sequence version replaced gi:22038291.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht>
ml.

FEATURES

source	Location/Qualifiers	Pred. No.:	Length:	205283
repeat_region	1. .205283	Score:	2891.50	Matches: 1166
repeat_region	/organism="Homo sapiens"	Percent Similarity:	20.05%	Conservative: 56
STS	/mol_type="genomic DNA"	Best Local Similarity:	19.13%	Mismatches: 271
repeat_region	/db_xref="taxon:9606"	Query Match:	21.88%	Indels: 4610
repeat_region	/chromosome="12"	DB:	9	Gaps: 49
repeat_region	/clone="RP11-408I18"	US-09-522-753-5 (1-2517) x AC073916 (1-205283)		
repeat_region	/complement(1. .116)	QY	752	ProSerProHisThrGluAlaAala-----LysAspThrGlyGln 764
repeat_region	/complement(185. .545)	Db	65549	CCAGCACCGTACTGACAAACACCGCAGCTCTCTCATTTGGCCATGAAGAGCCATGGCCAG 65490
repeat_region	/rpt_family="WLTA0"	QY	765	AsnGlyProLysProAlaThrLeuGlyAlaAspGlyProProGlyProProThr 784
repeat_region	482. .856	Db	65489	AGTGGCCCTCGCCCACTGTGTCCAGGCTCTTGCTGGGAGCCCCATCTCTCCTCTCT 65430
repeat_region	/standard_name="D12S820"	QY	784	-----
repeat_region	548. .569	Db	65429	CTAGGCTCTGGGTTCAGAAACGAGGAGACCTCCAGGAAGGAGTTAAGGAATCGAGTG 65370
repeat_region	/rpt_family="AT_rich"	QY	785	ProProArgArgThrSerArgAlaProlle-GluPro-----
repeat_region	710. .845	Db	65369	CCGGAAAGAGAGAAATTCCTGGCAGCCTAGGCGACCCAGGGGTGGAGATGAAGTGCT 65310
repeat_region	/rpt_family="(TCCA)n"	QY	797	-----ThrProAlaSerGluAlaThrGlyAlaProThrProPro-ProA 811
repeat_region	/complement(846. .980)	Db	65309	AATGGGGCGCTCTCTCAGCACTGCAGCTGGAGCGCCGGAATTCCTCTCTCCATCCAC 65250
repeat_region	/rpt_family="L2"	QY	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	1320. .1681	Db	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	/rpt_family="L2"	QY	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	2124. .2158	Db	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	/rpt_family="AT_rich"	QY	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	2327. .2353	Db	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	/rpt_family="(T)n"	QY	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	/complement(2835. .2914)	Db	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	/rpt_family="MER58B"	QY	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	3015. .3377	Db	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	/rpt_family="L1MC4"	QY	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	/complement(3384. .3882)	Db	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831
repeat_region	/rpt_family="LTRA0a"	QY	811	laProProSerProSerAlaProProValValProLysGluGluLysGluGluGluT 831

65249	TTCGGCTGTGCCCCGAGCCCCCTCCCCAGGCGCTG-----G	65211
Qy	831 hrAlaAlaAlaProProValGluGluGluGluGlnLysProAlaAlaGluGluL 851	
Db	65213 GAGGTGGAGGTGGCACCGGTGTG-GCTTAGGAACATAATGCATCTCCTGCTGCCACAGAGA 65155	
Qy	851 euAlaValAspThrGlyLysAlaGluGluProValLysSerGlu-CyethrGluGluAla 870	
Db	65154 TAGCCTTTGGAGACAGCG-----CTGCAGCTGTGTCTTGCGGTGCC 65116	
Qy	871 GluGluGluProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGluAla 890	
Db	65115 AGTCTATGCCCT-----GGTGGCCCTGGACCGAGTGGCCCTGGGGGTGGCG----- 65071	
Qy	891 LeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSer--- 909	
Db	65070 -----GGAGCCTGGGAGGGCTGGTGGTGGGGTGGTAGTCNAGAG 65032	
Qy	910 -----GlyAlaProGlnAspSerAspSer 917	
Db	65031 CTTGTCTTGAGAGTCACTGGGTAGAGTCCCACTTGGGACCCACAGACGAGTGCCTGAGC 64972	
Qy	918 SerAlaThrCysSerAlaAspGluValAspGluAla---GluGlyGlyAspLysAsnArg 936	
Db	64971 CTTTATAGGCCTTTCACGCGTATCGTCTTCATCATGGGTTTCAGTCGGGGCCTTTAAACTCT 64912	
Qy	937 LeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProAlaAsnAlaSer 956	
Db	64911 CGTCTGCTCCTGGCGAGGT----- 64891	
Qy	957 ProGlnLysProLeuAspLeuLysGlnLeuLysGlnArg----- 969	
Db	64890 -----AGGCAGTCAGCAGCGCGCAGGTGTGAGACTGTGAGACTGTAGGGAG 64852	
Qy	970 -----AlaAlaAlaIleProProIleGlnValThrLysValHis 982	
Db	64851 TGGGGAGGAGCTGTAGGGAAATGGGGAGGACTGTGGTGCCTGTCTCAGCC-----GTCCAC 64798	
Qy	983 GluProProArgGluAspAlaAlaProThrLysProAlaProAlaProAlaProProPro 1002	
Db	64797 TCCCT-----CGGGCCACCGGTGACCCAGGTTTCTCATCTCCGTTTATCT 64750	
Qy	1003 Gln-----AenLeuGlnPro-----GluSerAspAlaProGlnGlnPro 1015	
Db	64749 GAAATCTCCGATGTTTAAACATCGCGGATTAATTTGGAAACGTTTTCTCTGAACAGCAACT 64690	
Qy	1016 GlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAlaPhe 1035	
Db	64689 AGTACCT-CCTTTGGCAAC-----CCTGGAGTAGCTACGCGGCCGTG 64646	
Qy	1036 AlaAlaGluAlaGlnLysLeuProGlyAspProCys----- 1048	
Db	64645 GGCACACGAGCAATGTTGAAACCCGAGAGCCTGTCACTTCTCATCTCCCTCTCTGAG 64586	
Qy	1049 ---TripThrSerGlyLeuPro-----PheProValProPro 1059	
Db	64585 GGGTGGGG-CGGGGGCTCCGGGGCTCATTTCTGTAGTCTCTGGACTCGGCTCGCCTGGA 64527	
Qy	1060 ArgGluValIleLys-----AlaSerPro-----HisAlaProAspProSer 1073	
Db	64526 AGGAGAGCCCTGGCTAGATGGGCAAGCCGACCTTTACCTTCGGGGGGCCACCTCCGCTCT 64467	
Qy	1074 AlaPheSerTyrrAlaProProGly-----HisProLeuPro-LeuGlyLeuHisAspTh 1091	
Db	64466 GTTCACCTCTCTGCTCGGGGACAGTGAACCGGCCCATGTAGAGCTCTCTTTGGCCT 64407	
Qy	1091 rAlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSe 1111	
Db	64406 GGCAGGCCACCCCTGCCACCCCTCACCACTGCCTGCAAC----- 64367	
Qy	1111 rAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerVa 1131	
Db	64366 ----ACACACTCTCCGTGCACAGCAGCACTTGTGTGTCGACACTCATGTCATCATCGC 64311	

Qy	1131	lGln	-----	1132
Db	64310	ACAGGTTTCTGGGTGTGCACACGTTGATACATACAAAGCCGTGTACTGCTGCACCCCTGGT	64251	
Qy	1133	-----LeuHisValProTy	1137	
Db	64250	ACATGTGTGTACATGAGCCACCTTAGTTCTCAGCAGCCAGGCTCAGCTGCATGTGCCAC	64191	
Qy	1137	rSerGluHisAlaValAlaProValGlyProValThrMetGlyLeuPro	1153	
Db	64190	ATCCACTCTGTGCACACACAGACCCGTGCATGCATGCATGCTGCTCTGCTGCCACAC	64131	
Qy	1154	-----LeuProMetAspPro	1158	
Db	64130	GTCCACACACTGCTGATGTATCGGTGCACACGCGTCTCTGCTCTGCCGCGCACCTCCACT	64071	
Qy	1159	-----LysLysLeuAlaPro	1163	
Db	64070	CTGCTGTATCGGTGCACGCGTGTCTGCTGCTGCCACATCCACATCCACACTCTGCTGCTGATGT	64011	
Qy	1164	-----PheSerGlyValLysGlnGluLeuSerProArgG1	1176	
Db	64010	ATCGGTGCACGTGTCTCTGCTGCCACACATCCACACACTGCTGATGCTGCTCTGTCTC	63951	
Qy	1176	yGlnAla	1189	
Db	63950	ACGTGCATGTGTGCACCTTGTTCACGCCCCATGTCCGACCCCGTGGTGTGGACACAGA	63891	
Qy	1189	gGlnAlaSerValLeuArgGlyThrAlaLeu	1207	
Db	63890	CTCACGTGCTCATGTGTGCACAGGCACACCTTGTGTGCAGACAGACACCCCTGGAGGCC	63831	
Qy	1207	rIleThrLysGlyLeuProSerThrArgValProSerAspSerAlaIleThrTyArgG1	1227	
Db	63830	TAGGGTACAGGGTGCAGGACGCTGCCCTCGCATCTCCACCGTCGACACCCCTGGAAAAG	63771	
Qy	1227	ySerIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrIleThrArgIleI1	1247	
Db	63770	CTCTGG-----GTCTGCTGGCAGGCCCC-----CA	63744	
Qy	1247	eGlyLysAspSerProSerArgLeuAspArgGlyArgGluAspSer	1262	
Db	63743	GGGCTGCAGGCTGCCAGCTCCCTCTGGAGGGCTCGGCTGTGAGGCTTTGTGACGGGGCC	63684	
Qy	1263	-----LeuProLysGlyHisValIle	1270	
Db	63683	AGCATGAAGACACTGTGGCTCTGCTCTACCGGCTCTGCTGTCTGCTGCCACCCGCTG	63624	
Qy	1270	rGluGlyLysLysGlyHisValLeuSer	1285	
Db	63623	GTCTGGGCCGCTGCCAGCTCTCACCCAGTCTTTGTGTGGTTTCAGGTGGCATGTC	63564	
Qy	1285	rValThrGlnCysSerLysLysAlaSpGlyArgSerSerGlyProProHisGluThrAl	1305	
Db	63563	TGTGACCCAGTGTCTCAAGGAGGAGCGAGAGAGCTCAGGACCCCCCATGAGACGGC	63504	
Qy	1305	aAlaProLysArgThrTyAspMetMetGluGlyArgValGlyArgAlaIleSerSerAl	1325	
Db	63503	CGCCCCAAGCGCACCTATGACATGATGAGGGCGCGTGGGCAGAGCATCTCTCTCAGC	63444	
Qy	1325	aSerIleGlu	1328	
Db	63443	CAGCATCGAAGGTATGAGGAGGAGAGACTTCATCTCTCGGTGCCCTCTGTTGGGGCGGT	63384	
Qy	1328	-----	1328	
Db	63383	GGGGGATGGCTGACCCCGTTTTACAGATGGGGAACCGAGGCTGGGCTTTTCTGAGGCTC	63324	
Qy	1328	-----	1328	
Db	63323	CATCTGAGGTAGCGGAGGGAACCTTCCCGTGTGGTCTCTACTCTCCACCATCATCTGGGG	63264	

QY	1573	-----	1573	-----	Db	57805	AATACAAAAA	TAGCCGGGTGTGGCGGCGCCTGTAGTCCAGCTACTCGGGAGGCT	57746
Db	58885	GAAAGAAAAATTTCTCTGAGCGTGTTCACCTGTGAATTTCTCATTTGTTGCTAGTTTA	58826	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57745	GAGCGAGAGAAATGGCGTGAAACCCGGGAGCGGAGCTTCAGGTAGACCAAGATCGCGCCA	57686	
Db	58825	TCACCTGTCTCTCGCATTTGAATGTTCAGCTTGTGAGGCGCTGGGATTTCTGTTTCGTTCACT	58766	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57685	CTGCATCTCAGCCCTGGGCGACAGAGCGAGACTCTCTTCTTAAATAATAAAGAAATTAATA	57626	
Db	58765	GGGGTGACCCCAAGTTCTCAACAATGCTTTGCCACGTAGTAGAGCTGCATCAATATTTT	58706	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57625	AAAAAGAAATAAAAAAAGTTGCATCCCTTTTGGAGTGTAAATCTGCATTTGGATGTCCTAT	57566	
Db	58705	TTAATTGATTGAGTGAGTGAATGATGAAGAAATGAATTTTTTAAAAAACTATAACAAAA	58646	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57565	GTTTGGGACAACCTTTGATGCAAAAAGCATCTTCGTAGAAGTCACCTCTTGTGTCTGG	57506	
Db	58645	AGCAAAATGAGTCAGTGAGCAAAAAGTGAACCTAAGGCAATGAAGAAATGAAGGAGTGAATG	58586	-----	QY	1644	-----	-----	1644
QY	1573	-----	1573	-----	Db	57505	CGTGATGTTTTCTGTCTGTCCGACGCTCAGTTCTGGTTTGTGTGTTGGGCGACCAACAT	57446	
Db	58585	AAGAGACCTGGTCTCTTGGGATCCGAGGTCCCTATCTCTCAACAACCTCCCGTAAATGCC	58526	-----	QY	1644	-----	-----	1644
QY	1574	-----	-----	-----	Db	57445	GTAGTGGGAGAGCTGTCGGGTGCAGAAAGTAGGGGCATCCAGACAGGTGGAGCGACA	57386	
Db	58525	AGCCCCAGAGCCCGATGCATCCACCTTGCCCGTCCACAGCGCAGCCCTTCTCGTCCAGCAAG	58466	-----	QY	1645	-----	-----	1645
QY	1579	AlaSerGlnAspArgLysLeuThrSerThrProArgGluLeuAlaLysSerProHisSer	1598	-----	Db	57385	CCATCAGCGCTAGGTATGGCTGGCCCTCACATGAGCTCCCTCTGCCCCGCGAGCGCTGCC	57326	
Db	58465	GCATCCAGAGCCGAAAGCTGACGTGCAGCGCTCGTGAGATGCCAAGTCCCGCCGACAGC	58406	-----	QY	1649	TyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyr	1668	
QY	1599	ThrValProGluHisProHisProHisProHisProHisProHisProHisProHisProHis	1618	-----	Db	57325	TACTACTGCCCCGACACCTTGGCCCCCAACCCCACTTACCCGACCTGTACCCACCTAC	57266	
Db	58405	ACCGTGCCCGAGCACCAACCCCACTCCGCTTATGAGCACCTGCTTCGGGGCGTG	58346	-----	QY	1669	LeuLeuArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrLeuLeuAsnAsp	1688	
QY	1619	SerGlyValAspLeuTyrArgSerHisLeuProLeuAlaPheAspProThrSerLeuPro	1638	-----	Db	57265	CTCATCCGCGGCTACCCGACACGCGCGCTGAGAACCGGACAGCATCATCAATGAC	57206	
Db	58345	AGTGGCGTGACCTTGATTCAGGCCACATCCCTCTGGCTTTCAGCCCACTCCATACCC	58286	-----	QY	1689	TyrLeuThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAsp	1708	
QY	1639	ArgGlyLeuProLeuAsp-----	1644	-----	Db	57205	TACATCACCTCGCAGCAGATGCACCAACGCGCGCCACCGCCATGCCCCCAGCAGCTGAT	57146	
Db	58285	CGCGGCATCCCTCTGAGCAGCGAGTGATTGCCCTGGGGCTCCACAGAACCTTGACGTGTC	58226	-----	QY	1709	MetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyPro	1728	
QY	1644	-----	1644	-----	Db	57145	ATGCTGAGGGGCTCTCGCCCCCGAGTCTCTCGCTGGCACTCAACTACGCTGGGGTCCC	57086	
Db	58225	TGAACAGGGCCACGGACCTCATCAGTGTTCCTCAGGACTCCTTAGGCATCAACTGTCA	58166	-----	QY	1729	Arg-----	-----	1729
QY	1644	-----	1644	-----	Db	57085	CGAGGTGAGTGGGTGGGCGACACCCCTCCGCTGGGTTTGGCCTTATTCCCAAGGACATG	57026	
Db	58165	GTTTCCCTCGATGGCGAAACTGAGGCTCGGGATTGGAAAGACCAACAGTGAATCATG	58106	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	57025	GGCGTGCCCTGTGGGCTCGCGGAGCAGCTAGACCTGTGTCACTTGTGGGTCACTTGT	56966	
Db	58105	AGCTTAGGTTGGAGCAGAAATTTCTCTTAGTAGTTTGCAGGACATGTGGGGTTAAACATTT	58046	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	56965	GTGAACGGACCTCAGTGGGTGGGCTGGGGTTGTGCGTGTGTGGGTCTGTGGTGGCATCT	56906	
Db	58045	CAGTGGTGTCTTTTCCGGCAGGACTTATCAGTGCCTTTTAGCAATGCAAGAGGTATAGAA	57986	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	56905	GGTAGGTGAGTGACACAGCGTGTGGCTCTGGCTGCATCCCTCAGTGGGTGTGCGTGCATCT	56846	
Db	57985	GAGACTTGAGTATATCATTTTCAAAATAGACATGATCTGAAAGTCTTTTAAAGTT	57926	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	56845	GTGTATATCTTCTTAGGATACAGGGGCTCAGAGAGTTTAAAGATCAAAATGTGGCGGGCAC	56786	
Db	57925	GCCGGCAGCGTGTCTCACACCTGTAATCCAGCACTTTTGGAGCGCGAGCGCGAT	57866	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----	Db	56785	AGTGGCTCATGCTGTAAATCCAGCACTTGGGAGGGCGGAGGAGGTGGATACAAGGTC	56726	
Db	57865	CACAAGTTCAGGAGATAGAGACCATCTCTGGCTAACCGGTGAACCCCGTCCCTACTAAA	57806	-----	QY	1729	-----	-----	1729
QY	1644	-----	1644	-----					

Db 56725 AGAAGTTGAGACCACTGTGACCAACATGGTGAACCGGTCTCTCTTAAATAACAAAA 56666
Qy 1729 -----
Db 56665 TTAGCCAGGCATGGTATGCGCACTCTGTAGTCCAGCTACTCTACTAGGCTGAGGCAAGAG 56606
Qy 1729 -----
Db 56605 AATCACTTGAACCCAGGAGGTGGAGTTACAGTGAGTGGAGATTTTACCATTTGCATCCA 56546
Qy 1729 -----
Db 56545 GCCTGGGCAACAGGGCAAGACTGTGTCTCGMAAAAAAAAAAAGATGAAATGTGAGGCT 56486
Qy 1729 -----
Db 56485 GTTTGGAGTTGTTTCCTTTGCTTTGTAACAGCCCAAGCTGCTTTTGGTGCACACGTTT 56426
Qy 1729 -----
Db 56425 CAGGGCCATCTCAGAAATGCTTCTGGAATAACCAAGTTCTTAGCTGGGGCTCAGCTGGAA 56366
Qy 1729 -----
Db 56365 AAGCTGAAGTCACACTTAAGTATTTTGAACAGTGAGGATGAATACAGGGAATGGATTGT 56306
Qy 1729 -----
Db 56305 GTAGCGTACAGGCTGAAGGGGCACAGAGGCTTGAGATGGGAACCAAGTGAGGGCAGCT 56246
Qy 1729 -----
Db 56245 GCAGGAGATGCCCGGCTCGGGCTTGGGAGCAGAAAGAGAGGTGTACTCCGAGAACCTG 56186
Qy 1729 -----
Db 56185 AGCATTCAGAAAAAGGTTCCATGGCTGGTGTGGAGCCGAGGAGGAGTGCCCTGCCACC 56126
Qy 1729 -----
Db 56125 AGCTCTGCTCCAGAGTGTGTGCGGTCTCTCCAGAGGGTGATCTGGCGGTGGG 56066
Qy 1730 -----GlyIle 1731
Db 56065 CAGCTGGCCTCTCTCTCCTGTGGTACAGCCCTGGCCCAACACCTCCCGCAGGCATC 56006
Qy 1732 IleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrPro 1751
Db 56005 ATCGACCTGTCCCAAGTGCACACCTCTGCTGTGTCGTCGCCCGACACAGGCACCCCA 55946
Qy 1752 AlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg 1771
Db 55945 GCCACCCCATGACCCCTTGCTACTCTCCCAACCGGCCCGCCAGCCCTTCAGCAGCCGC 55886
Qy 1772 HisSerSerProLeuSerPro- 1779
Db 55885 CACAGCAGCTCCCCACTCTC-CCCAGTAGCGCCACTGCCAGTCTGGGTGGGGACCCC 55827
Qy 1779 -----
Db 55826 GGCATCCATGGAGCGGCTGGGGATGGGGGAGAGAGCCCTCTCTCTTTCCCAACC 55767
Qy 1779 -----
Db 55766 CAGAAGCAAAAGCAGGCTCTTCTTCGGCCCTGGGGCTGAGTCTCTGGCCTTTGGGTTTC 55707
Qy 1780 ---GlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArg 1798
Db 55706 CTAGGAGGTCACACACTTGACAAACCAACCAACCAACGCTCTCTGTCGAGCGGGACGA 55647
Qy 1799 AspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThr 1818
Db 55646 GACCGGGATCCAGAGCGGGACCGGGATCGGGAGCGGGAAAGTCCATCTCTCACTCCATCCACC 55587

Qy 1819 ThrThrValGluHisAlaProlleThrArgProGly----- 1830
Db 55586 ACACCGTGGAGCAGCAGCCATCTGGAGACCTGGTAGGGCATCAGAGCCCCACCCCCC 55527
Qy 1831 -----ThrGluGlnSer 1834
Db 55526 GTCTCCGGACTCTTGTGGGCCCAAGAGGCTCCCTCTGTGATGCCACTGACTG-TCA 55468
Qy 1835 SerGly-----SerSerGlySerSerGlyGlyGlyGlyGlySerSerSerArgPro 1851
Db 55467 CCAGGTACAGAGCAGAGCAGCGCAGCAGCGCGGGGTGGGGCAGCAGCAGCGGCC 55408
Qy 1852 AlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeu 1871
Db 55407 GCCTCCCACTCCATGCCACCAAGCACTCGCCCATCTCTCCCTCGACCCAGATGCCCTC 55348
Qy 1872 GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGlu 1891
Db 55347 CAGCAGAGACCCAGTGTGCTTCAACACAGGCAATGAAGGTATCATCACCGCTGTGGAG 55288
Qy 1892 ProSerLysProThrValLeu----- 1898
Db 55287 CCAGACGCCCAACCGTCTCT-GAGGTGGGCCAGGTTGGCATGGGGAGGGGGCGGCAGG 55229
Qy 1898 ----- 1898
Db 55228 TGGATGGGTGTCAGTAGGAGGATAGCAGATAAGAGGATGCTTGGTGGGAGGTATATGG 55169
Qy 1898 ----- 1898
Db 55168 GAGGTGGGTGGGTGGGCAGATGTGTGGGGGTAGAAAGATAGACTATGTGTGAGTAGTGA 55109
Qy 1898 ----- 1898
Db 55108 TGGGTGGGTGTCGGGTAGGTGGGTGTGTGGGTCACTGGGAAGTTAGGTAGTCTGGTG 55049
Qy 1898 ----- 1898
Db 55048 GGTGGATGGATGGATGGGGGGGTGTCGGGTAGGTGGGTGGGTGGGTGGGTGGGTGGAA 54989
Qy 1898 ----- 1898
Db 54988 AGATAGATGGCTGGGTAGCTGGTTGGATGGGTGGATGATCAGGTTTGGTAGGTAGGTGG 54929
Qy 1898 ----- 1898
Db 54928 GTGGGTGAAGAGATAGATGGCCCGGTTAGCCAGATAGGTGGTTAGGTGGGTAGGT 54869
Qy 1898 ----- 1898
Db 54868 GGCTCAGTGGGAGGATAGATGGTTAGTTAGCTGGGTGGGTGGATGGATGGATGGGTGGG 54809
Qy 1898 ----- 1898
Db 54808 TGATAGGGTTAGGTAGGTAGGTGGGTGGGTGGATAGATAGCTGGGTAGCTGGGTG 54749
Qy 1898 ----- 1898
Db 54748 GATGGATGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 54689
Qy 1898 ----- 1898
Db 54688 AAGTTAGCTGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 54629
Qy 1898 ----- 1898
Db 54628 AGGTGGCTAAGAGGATAAATGACTGAGTTAGCCAGATGGATGGATGGATGGATGGATGG 54569
Qy 1898 ----- 1898
Db 54568 TGGATGGATGGGCGGCTGGTTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGATGA 54509

Db	52348	CCTACCTCATGAGCGCGTCTTGCTGCCCAAGAGAGCGCCCCCGGGTGC	CGCCGACAGAG	52289
Qy	1950	ArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeu	1969	
Db	52288	CGGCCCGAGCAGACACCGGCATGCGCTTCTCGCAAGCGCCCGCGCTCCGGGCTG	52229	
Qy	1970	GluProAlaSerSerProSerLysGlySerGluProArgProGluProValProValSer	1989	
Db	52228	GAGCCCGCGCTCTCTCCCGCAGCAGGGCTCGGAGCGCCCGCGCTAGTGCCTCTCTCT	52169	
Qy	1990	GlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro	2009	
Db	52168	GGCCACGCCACCATCGCCCGCACCCCTCGCAAGAACCTCGCACCTCCACCGCCGCG	52109	
Qy	2010	AspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLys	2029	
Db	52108	GACCCGCGGCGCCACCTGCTCGCTCGGCTCGGACCGCGCACCGGGAAGACTCAAAATAA	52049	
Qy	2030	ProPheSerIleGlnGluLeuLysArgSer	2040	
Db	52048	CCCTTTTCCATCCAGGAACGTGAACTCGTTCTCTGGGTAAACACCCCTGACAGCGGCC	51989	
Qy	2040	-----	2040	
Db	51988	ACCTTCATAGACGCGATTATCATCGTCAAAATGCTCACGATAAAGGGCGCAGAAAGA	51929	
Qy	2040	-----	2040	
Db	51928	GGTGGCTGGCCACGGCTCCCTCGCGATGGTAAGACTTCCGGCCCGCACCCACCCGCT	51869	
Qy	2040	-----	2040	
Db	51868	CTCGTGTCCAAAGATATTTTCAGATCTCTGCTTTTACTTTTGGCCCCGTTTTTTTGT	51809	
Qy	2040	-----	2040	
Db	51808	TGTTGGTTTTGGTATTTTGTGTAAGCCCATCGTCTCTCGCGGTTTTTGACGCGCTGAC	51749	
Qy	2040	-----	2040	
Db	51748	GACTACTCGGCGCGCGCTGCGCTCTGGTTGGGGGGGCTCATCTTTGCACATCATTT	51689	
Qy	2040	-----	2040	
Db	51688	TACCATGGTTTTTTTTTTTGGATTTTGTGCTTTTCTTTTATGAATGATCTGTGAT	51629	
Qy	2040	-----	2040	
Db	51628	TCTGACTTCGACTGCGGCCCATCTCCCTCTTTTGGCCTGTGTCCAGGGAGCAGGATGG	51569	
Qy	2040	-----	2040	
Db	51568	GGCTGGGGAGGGCTCGGGGCTACGCCGCCACCTGCGCGGTGCCTGGATGCTGCGGACT	51509	
Qy	2040	-----	2040	
Db	51508	GGGGGAAGTGGAGCAGCGGTGCAAGAGAAGCTGAGGGGGGCGAGGACCTCGCGTGT	51449	
Qy	2040	-----	2040	
Db	51448	CGAGGAGAGTGGGTCTGGCTCTTGCATCTTGCCCTGTCCCGAGCCCTGTACCCGAGA	51389	
Qy	2040	-----	2040	
Db	51388	AAAAGGGAGCCCTCTGCTCTGACCCCTGCTTGGCCCTAGTTTCAATGGCTCTCTCTG	51329	
Qy	2040	-----	2040	
Db	51328	TTGGACTGGGATGGCCGAGGCTATAGCCAGCGGGGGCCCGGGGACCCAGGGTCACTCC	51269	
Qy	2040	-----	2040	
Db	51268	CAGCCACACCCCACTTCTCACTCCGGCCACACACATCTCTTCCCGCAGAGACCCATGCTGC	51209	

Qy	2040	-----	2040	
Db	51208	CCCATCTCAGCTGGGCTCGCCCGGCTCCACACCTGAACCATCTCTGTCCCTTCTT	51149	
Qy	2040	-----	2040	
Db	51148	GCCTAATCTCTCTGTGTCTCCCTCTCTGTGTCTCTGTGTCCCGGGCTCTGCATCTCTC	51089	
Qy	2040	-----	2040	
Db	51088	TCCACCTCTCCCTTGGCTCCCTGTCTCTCCCCAACACCCCTCTCTGCTTACTGTCTTT	51029	
Qy	2040	-----	2040	
Db	51028	GGGAGCCCAAAACCTACCCCTAGCTTGGGTTCCCTTTGACCCCCCGGGGTCCACGCCAG	50969	
Qy	2040	-----	2040	
Db	50968	CTGGAGGGCAGCCCTCGCCCTCGGGCTCCGAAACCTGGGCCCGGTCCTGACTCTGCA	50909	
Qy	2041	-----LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProVa	2056	
Db	50908	CCCCCGCGCTGCGCTAGGTTACACGCGCAGAGCTACAGCCCCGAAGGGGTGAGCCCGT	50849	
Qy	2056	lSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluG1	2076	
Db	50848	CAGCCCTGTGAGCTCACCCAGTCTGACCCACGACAAGGGCTCCCCCAAGCACCTGGGA	50789	
Qy	2076	uLeuAspLysSerHisLeuGluGlyLeuArgProLysGlnPro	2091	
Db	50788	GCTCGACAAGAGCCACCTGGAGGGGAGCTGCGGCCCAAGCAGCAGTACGCCCCACCC	50729	
Qy	2091	-----	2091	
Db	50728	AGTACCAGGCCCCGGAAGCCCTGCACTGAGGACCCCTCAAGSCCCCATCATGACAGATA	50669	
Qy	2091	-----	2091	
Db	50668	GGRAAACAGAGGTGCCTAAAGGCCAAGGAATTGGCTGGATCATGAGGCTCAAAGGCATGG	50609	
Qy	2091	-----	2091	
Db	50608	GGCTGGGATTGGGCCCGCAGCAGCCCTGGGGCCAGACAGACCCCTCACCAGCTCTGCTC	50549	
Qy	2092	-----GlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProL	2109	
Db	50548	TGCTGCAGGCCCGCGTGAAGCTTGGCGGGAGGCGCCACCTCCACACCTCGCGGCGC	50489	
Qy	2109	euProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyH	2129	
Db	50488	TGCTGAGAGCCAGCCCTCTGTCAGCCCGTGTCTCCAGACGCGCCCGAGGGGTCAAAGGTC	50429	
Qy	2129	isGlnArgValValThrLeuAlaGlnHisIleSer	2140	
Db	50428	ACCAGGGGTGGTCACTGCGCCAGCACATCATGTAACCTAGTAACTACGCGTTCTCTGCTGCTGC	50369	
Qy	2140	-----	2140	
Db	50368	TTGTCACTTTTGACCTGGGGGCAACAGGCTTGGAGAGGGGATGGGGAACCCACAGCCC	50309	
Qy	2140	-----	2140	
Db	50308	TTCTGTCTGGCGGGTGGCTGGGGATCCAGGGCATGGCGCTGGGGGATCCAGGGCGT	50249	
Qy	2140	-----	2140	
Db	50248	GGGTGAGGGTGAATCCAAAGCCCGCAGCACCGGCACCATCACCGCCCCCTAATCCATG	50189	
Qy	2140	-----	2140	
Db	50188	GGAGGAGCCTGTGATGCGAGCCGATGGCATTTTCAOGGGCAATGAGGCCCTTCTGTGTGGC	50129	

QY	2140	-----	2140	Db	49048	ACTCCCACTCCACACCCCGCCGCCACCCCACTGGTGAACCGTGGGCGCTCTGGGGGT	48989
Db	50128	CCAGTTTCTCAGTGTCTATGGGCTGGTCTCATCAGCCATCTGCCAATCAGCTTGGGA	50069	QY	2204	-----	2204
QY	2140	-----	2140	Db	48988	CAAAAGAGAAAGAGAGGGAGGGCCCTGAGCTCCAGGGTGAAGAGGCGGTTTGGTGGGG	48929
Db	50068	CCGTGACACAGACCCCACTCCATGCATGCACACTGGGACACGAGGCCACGAGGGTGGCGGG	50009	QY	2204	-----	2204
QY	2140	-----	2140	Db	48928	GCGGGGGGGTAAATGTGTGTGCGAGACAGACATGSCAAGACAGCAGGACATCTTTGGGG	48869
Db	50008	CAGTCCACAGTCAACCAGGAAGCTGGCCCCACCCAGGATTCCTGCCCGAGCTCCGTCTA	49949	QY	2204	-----	2204
QY	2140	-----	2140	Db	48868	GGCAGTGTGTAGCTGGCACTGGGCTACCAACAGAAATTCAGACGAGGGTCAATAAACATG	48809
Db	49948	GCCCCCCCCACCCCAAGAGTTCTGTCAAGAGAGTGTGCTGACTCTGGGCCCCCCC	49889	QY	2204	-----	2204
QY	2141	-----	2141	Db	48808	TGCCCCATGGCCAGATCTAGCCCCGGGCCCTCTGTTTGTACAATTCATGAATCAAAAAA	48749
Db	49888	ACTTGCCTGCAGGAGGTCAACACAGGACTACACCCGGCACCAACACAGCAGCTCAGC	49829	QY	2204	-----	2204
QY	2157	AlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeu	2176	Db	48748	TGATTTTACATTTTAAAGGGTTGTTTAAAAAATAAATAATGATAACGATACATGCC	48689
Db	49828	GCACCCCTGCCCGCCCCCTCTACTCTCCCTGGGGCCAGCTGCCCGCTCTGGACCTC	49769	QY	2204	-----	2204
QY	2177	ArgArgProProSerAspLeuTyrLeuProProAspHisGlyAlaProAlaArgGly	2196	Db	48688	AGAGATTACTTGTGTGTCGAAAAATGCCTAAACCGTTTATGATTTGGCCACGTACAGGAAA	48629
Db	49768	CGCGCCCAACCCAGTGACCTTACCTCCCGCCCCCGGACCATGCTGCCCGGCCGTGGC	49709	QY	2204	-----	2204
QY	2197	SerProHisSerGluGlyGlyLys	2204	Db	48628	GCGTGTGGGCCCCGTGTTTAGAGTGAGGAGGGTGCGCCCTCAGTAGGGAGACCTCT	48569
Db	49708	TCCCCCACACGCAAGGGGCAAGAGGTGAGCGAGGATGGGGTGCCTCTTGCTGTGTC	49649	QY	2205	-----	2205
QY	2204	-----	2204	Db	48568	GACCACATCTGGGGCCCTTCTCCATCCAGGTCTCCAGAGCCAAACAGACGTCGTCTT	48509
Db	49648	GCGGGGGAGGCTTCGTGAGTGCCTGCCTCAGAGCTGGGTCTGAGTCAAGTGCAT	49589	QY	2215	uGlyGlyGlyGluAspGlyIleGluProValSerProGluGlyMetThrGluProGlu	2235
QY	2204	-----	2204	Db	48508	GGGTGTGTGTGAGGACGGTATTGAACCTGTGTCCCCACCGAGGGCATGACGAGCCAG	48449
Db	49588	TGAGGGCCGCTGGTCCATTGCACAGGCAAGAAATGGAGGCTCCAGGAGATGGGGAGT	49529	QY	2235	yHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSe	2255
QY	2204	-----	2204	Db	48448	GCATCCCGAGTGTGTGTACCCGCTGTGTACCCGGATGGGGAAACAGACGAGGCCAG	48389
Db	49528	CACAGACGGATACATAACGAATAATAGCTGATGAAGTATACATGCTGTGCTCAGTGGTT	49469	QY	2255	-----	2255
QY	2204	-----	2204	Db	48388	-GTACTTCTGTGGGACATGCGTGCCTCCCGGGATGCTCTGGTATCCCTTGCCCATCTT	48330
Db	49468	CCAGGCACTCAGTGTATACCTTTACACACAGCCCCCAGAGACATAGGAGCATTTCTCC	49409	QY	2255	-----	2255
QY	2204	-----	2204	Db	48329	GGCCCCAGTCCACCGTGGTGCCATGTGGAGAGTGACAGGGCACAGGGCTCAGTGGGTG	48270
Db	49408	CCATTTCACAGATGAAGAACTAAGGCTTGAGAGTGAATTCACTTGCCCAAGGGGCA	49349	QY	2255	-----	2255
QY	2204	-----	2204	Db	48269	ACCTCAAGCTGCAAGCAGGTTTCCACAACTTTGGGGGTGTGATACATGCCCACTCC	48210
Db	49348	CAGTCAGTAGGTGACAGAGGGGATTTGTTCCAGGGCTGTATGATTAGAGAGCGCTTT	49289	QY	2255	-----	2255
QY	2204	-----	2204	Db	48209	TGGGGAGACCTCAGTGTGTAGCTTTCCATGCTTTGGGCTGGGATCTCAGCTGGACAGGCC	48150
Db	49288	CCACATGCACACCCACACATGCACACACATGCACACATGCACACCCGACATGCGC	49229	QY	2255	-----	2255
QY	2204	-----	2204	Db	48149	CCCGTTGACAGCCCTGTTGAACCTCTAAGAAATAATGAGTAGTGTGTGCGGCCAGAC	48090
Db	49228	ACCGGCACACACAGTGCACACCCACACATATGCGCGGCACACATGCACACCCGC	49169	QY	2255	-----	2255
QY	2204	-----	2204	Db	48089	CGGGGGCGATGGGTGAGTGGGACCTGAGAAGGAAGCTGGGCCCTGCCCCCTGGGAAG	48030
Db	49168	ACATGCGCGGCACACACATGCACACCCGACACATGCGCACACACATGCACAC	49109	QY	2255	-----	2255
QY	2204	-----	2204	Db	48029	GGCATGATCGGAACCAATTTAGTCCCTTGGGGCTCTCTTGAGAGGGTCAGGCTGGAGCA	47970
Db	49108	CCGCACACAGCGGCACCCGCACACCCCTGTAGACACAGGCACATGCATGCTCAGC	49049	QY	2255	-----	2255
QY	2204	-----	2204				

Db 47969 AGTGGTCAGAGCCAGCCACACGACGACGAGCGTCTCCCAAGACACCCCTAGCCCCAGG 47910
Qy 2255 ----- 2255
Db 47909 ACGGGTGAACCGCAGAGGAGATTTTCAGAGCCGCGTATCTTACAGGCGGGGATGCAG 47850
Qy 2255 ----- 2255
Db 47849 GCGTGGGGGGGTGAAGCTTGTCTTCAAAATGTCTAAGGCATCTCAGGTGGTGGTTCCTCC 47790
Qy 2255 ----- 2255
Db 47789 CATCATCAAGGCGATGCAAGCTCGGCACCAAGTGAAGTGTGAGTGTGTGATCTCA 47730
Qy 2256 ----ArgMetGlySerLysSerProGlyAsnThrSerGlnProAlaPheSerLy 2274
Db 47729 CAGCAGATGGCTCCAGTCTCAGGCAACACAGCCAGCCGCGCAGCTTCTTCAGCAA 47670
Qy 2274 sLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluLeuLeuLeuLysLe 2294
Db 47669 GCTGACCGAGAGCAACTCCGCGCATGGTCAAGTCCAGAGCAAGAGATCAACAAGAGCT 47610
Qy 2294 uAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluLlePh 2314
Db 47609 GAACACCCACACGCGGAATGAGCTGAATACAGTAAGGGCGCTGACGGCTCCCGGGGAAG 47550
Qy 2314 eAsnMetProAlaThrGly 2321
Db 47549 CATGGGGCCACAGTGGGGCGG 47528
RESULT 21
AF303586
LOCUS 3120 bp mRNA linear PRI 01-FEB-2003
DEFINITION Homo sapiens NCOR isoform b mRNA, complete cds.
ACCESSION AF303586
VERSION AF303586.1 GI:28190007
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3120)
Yu, L.
Direct Submission
Submitted (08-SEP-2000) Institute of Genetics, Fudan University,
No. 220 Handan Road, Shanghai 200433, P. R. China
Location/Qualifiers
1. 3120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
85-2829
/codon_start=1
/product="NCOR isoform b"
/protein_id="AA032942.1"
/db_xref="GI:28190008"
/translations="MSSVYPPNQAGAFSTQSRSPHSVKYTFPSTHHODPAFGKH
EAPSPITSGQPCDDONASPSKSLKEIOSMDRVDREIAKVQOILKLKKQOQLE
EAAKPEPEKPPVPPVQKHSIVQIYDENRKAEEAHKIFEGLPKVELFLYNQ
SDTVKHENIKTYQWVRKULIFKRNHARKOREQIKORYDLMEAMEKKVDRIEN
NPRKAKESKTRYEYKQFEIRKOREQORFQVQORAGLSATIASEHEISIID
GLSEQNSKQMQQLSIPPMFDQORVKFINMGLMEDPMKVYKALVRNRYKRRGR
EKBIKDKFIQHPNFKGLIASYLERKSVPCVLYLYLTKNENYKALVRNRYKRRGR
NQOIAPOSERKEVEEKDEKTEKKEEKDEKDEKDEKDEKDEKDEKDEKDEKDE
ETEREQATPRGRKANSQGRKRITRSMTEAASAAAAAATPEPPPLPPPEP
ISTEPVETSEEMEVAKGLVHGRNAAIAKVMGTSEAKQCNFYNYKRRNL
DNLLQHKQTSKPREEDVQCSVASTVSAQEDIEASNEENPEDESEAGNS
DTESAPKSPVEAPKSESDPENATGTEPAVELEPTTETAPSPSLAVPSTPA
EDSVETQVNDSTI SAETAQMDVDQOEHSAEESVCDPPATKADSVDEVVPENHA
SKYEGTQNDLDRASEKVEPRDELVVAQIQAORPEQSDNDSDATCSADEVDG
EPRQRMPMDSKPSLLNPTGSLVSPKLPNFELDLPQLQRAAVIPMWVSCPCNIP
IGTPVSGYALYQRHIAKHESALLERQORQEQIIDLCEKRSSTSPCGTSKSPNREWSR
"

CDS

polyA_signal 3087.3092
ORIGIN
Alignment Scores:
Pred. No.: 1,05e-36 Length: 3120
Score: 2187.00 Matches: 500
Percent Similarity: 61.44% Conservative: 115
Best Local Similarity: 49.95% Mismatches: 216
Query Match: 16.55% Indels: 170
DB: 9 Gaps: 30
US-09-522-753-5 (1-2517) x AF303586 (1-3120)
Qy 137 LysAspArgSerLeuThrGlyLysLeuGlu----ProValSerProProSerProHis 155
Db 190 CAGGATCCAGCATTCGGAGGCAAAACATGAAGCTCCATCTCTCCAAATTCGGGGCAACCA 249
Qy 156 ThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuLeuGlnAsn 175
Db 250 TGTGGAGATGATCAAAATGCTTCACCTTCAAACTCTCAAGGAGAGATTAATACAGAGT 309
Qy 176 MetAspArgValAspArgGluLeuThrMetValGluGlnGlnIleSerLysLeuLysLys 195
Db 310 ATGGATCGTGTAGATCGAGAAATTCGAAAAGTAGAACAGCAGATCCTTAACCTGAAAAG 369
Qy 196 LysGlnGlnGlnLeuGluGluAlaAlaLysProProGluProGluLysProValSer 215
Db 370 AAACAACAACAGCTTGAAGAGAGGAGGAGCTAAACCTCTCAGCTGAGAGCCCGTGTCC 429
Qy 216 ProProProLeuGluSerLysHisArgSerLeuValGlnIleLeuLeuLeuLeuLeuArg 235
Db 430 CCTCTCTCTGGAGCAGAAACACCGCAGTATTGTCCAAATTTATTTATGATGAGAAATCGG 489
Qy 236 LysLysAlaGluAlaAlaHisArgIleLeuGluGluLeuGlyProGlnValGluLeuPro 255
Db 490 AAAAAAGCAGAAGAGCTCATAAATTTTAAAGTCTTGCCCAAAAGTTGAAGTGTGCA 549
Qy 256 LeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAla 275
Db 550 CTGTATAACCAAGCATCAGATACCAAGGTGTACCATCAGAAACATCAAGACAAACAGGTG 609
Qy 276 MetArgLysLysLeuLeuLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLys 295
Db 610 ATGAGGAAAAAATCATATTTTATTTTAAAGAGAAATCATGCAAGAAAAACAAAGGAA 669
Qy 296 GlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArg 315
Db 670 CAAAAATCTGCCAGCGCTTATGATCAGCTCATGAGGCATGGGAGAAAAAGTGGACAGA 729
Qy 316 IleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLys 335
Db 730 ATAGAAAATAATCTCGAGGAAAGCTAAAGAAAGCAAAACAGGGAATACTATGAAAAG 789
Qy 336 GlnPheProGluLeuArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGly 355
Db 790 CAGTTCCTCAAAAATTCGAAAACAAAGAGAAACAGCAAGAAAGATTTTTCAG--CCAGTTGGG 846
Qy 356 GlnArgGlySerGlyLeuSerMetSerAlaArgSerGluHisGluValSerGluIle 375
Db 847 CAGAGGGAGCTGGTCTTTTCAGCCCATTCGTAGGAGTGGAGCATGAGATTTCTGAAAT 906
Qy 376 IleAspGlyLeuSerGluGlnGluAsnLeuLysGlnMetArgGlnLeuAlaValIle 395
Db 907 ATTGATGGGCTCTCTGAGCAGGAAAGTAATGAGAAACAATCGGCAGCTCTCTGTGATT 966
Qy 396 ProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheLeuAsnMetAsnGlyLeu 415
Db 967 CCACCTATGATGTTGATGTCAGAACAAAGCAGTCAAGTTCATTAACATGAATGGGCTT 1026
Qy 416 MetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGln 435
Db 1027 ATGGAGACCCCTATGAAAGTGTATAAGATAGGACGTTTATGAATGTTTGGACGTGACCAT 1086

[illegible]

Db	2014	AGTCCTGAAATGCTACTTCTCGAGGAAACACAGAACCTCGCGTTGAGTTGAG-----	2067
Qy	804	GlyAlaProThrProProProAlaProProSerProSerProSerAlaProProVal	821
Db	2068	-----CCACACACGGAATGACCCAGTACATCTCCCTTACAGTTCCAAAGTACA	2121
Qy	822	ValProIysGluGluLysGluGluThrAlaAlaProProVal-----	837
Db	2122	AAACCCAGCTGAAGATGAAGATGTGGAGACCCAGGTGAATGACAGCATCAGTCTGAGACA	2181
Qy	838	-----GlulGluGluGluGluLys	844
Db	2182	GCAGAGCAGATGATGTAGATCAGCAGGAGCAGATGCTCAAGAGGGTTCTGTTGTGAT	2241
Qy	845	ProProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSer	864
Db	2242	CCGCCACCCGCTACCAAA-----GCTGACTCTGTGGACGTTCAAGTGAAGGTGCCAGAA	2295
Qy	865	GluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGlu	884
Db	2296	AACCATGCATCTAAAGTTGAAGTGATATACCAAGAAAGAGACTTGGATAGAGCC---	2352
Qy	885	AlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlySerGlyArgAlaThr	904
Db	2353	-----AGTGAGAAGGTGGAACCTAGAGATGAAGATTTGGTG	2388
Qy	905	ThrAlaLysSerSerGlyAla-----ProGln--AspSerAspSerSerAla	919
Db	2389	GTAGCTCAGCAATAAATGCCAAGGCCCGAGCCCGTCCAGCAATGATCCAGTGCC	2448
Qy	920	ThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer	939
Db	2449	ACGTGCAGCGCTGATGAG-----GATGTGGATGGAGCGCAGAGAGCGCAGAAATGTT	2502
Qy	940	Pro-----ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSer	956
Db	2503	CCTATGAGCTCAAAAGCCCTTCACTGTTAAACCCCACTGGATCTATCTCTGCTC---	2559
Qy	957	ProGlnLys-----ProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaPro	974
Db	2560	CCGTTAAACCAAAATCCACTGGATCTGCCACAGCTTCAGCATCGAGCTGCTGTTATCCCA	2619
Qy	975	Profile-----	976
Db	2620	CCAATGGTATCCTGCACCCCATGTAACATACCAATTGGAAACCCAGTGAGCGGTATGCT	2679
Qy	977	-----GlnValThrLysValHisGluProProArgGluAspAlaAlaProThr	992
Db	2680	CTCTACCGCGCACATTAAGCAATGCAATGAGTCAGCACTCTCTGGAG-----	2722
Qy	993	LysProAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAlaPro	1012
Db	2728	-----GAGCAGCGCGCAGACAGACACAGATAGATTGGAA	2763
Qy	1013	GlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProAlaAspLys	1033
Db	2764	TGTAGAAGTCTTCAAGTCCATGTGGACATCCAAAGAGTCCA-----AACAGA	2811
Qy	1033	GluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProCysTrpThrSerGly	1055
Db	2812	GAG-----TGG---GAAGGT	2822
Qy	1053	LeuProPheProValProProArgGluVal-----IleLysAlaSerProHis	1066
Db	2824	AGGTAGATAGACTATCCCAAGAGAACTTTTATGAAGGGAAATAAAGCTAAACCCAAA	2888
Qy	1069	Ala	1069
Db	2884	GCA	2886

RESULT 22
BC054296

LOCUS BC054296 1917 bp mRNA linear VRT 03--JUL-2003
 DEFINITION Xenopus laevis cDNA clone IMAGE:6879542, partial cds.
 ACCESSION BC054296
 VERSION BC054296.1 GI:32450290
 KEYWORDS
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 xenopodinae; Xenopus.
 1 (bases 1 to 1917)
 Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W.
 and Richardson, P.
 TITLE Genetic and genomic tools for Xenopus research: The NIH Xenopus
 Initiative
 Dev. Dyn. 225 (4), 384-391 (2002)
 JOURNAL 22341132
 MEDLINE 12454917
 PUBMED
 2 (bases 1 to 1917)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Vallalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 TITLE Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 JOURNAL 22388257
 MEDLINE 12477932
 PUBMED
 3 (bases 1 to 1917)
 Klein, S. and Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUN-2003) National Institutes of Health, Xenopus Gene
 Collection (XGC), National Institute of Child Health and Human
 Development, 6100 Executive Boulevard, Room 4801, Rockville, MD
 20892-7510, USA
 NIH-MGC Project
 CONTACT: XGC help desk
 Email: cgabs-x@mail.nih.gov
 Tissue Procurement: Drs. Donald Brown and Liqian Cai
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 BC Cancer Agency, Vancouver, BC, Canada
 info@bcgc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Lilias Prabhu, Parvaneh Saedi, Jacqueline
 Schein, Duane Smalios, Michael Smith, Lorraine Spence, Jeff Scott,
 Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 49 Row: 1 Column: 16.
 Location/Qualifiers
 1..1917
 /organism="Xenopus laevis"

/mol_type="mRNA"
 /db_xref="IMAGE:6879542"
 /clone="IMAGE:6879542"
 /tissue_type="Whole, tadpole, metamorphosis, stage 53"
 /clone_lib="NICHDP_XGC_Tad1"
 /lab_host="DH10B"
 /notes="Vector: pDNR-LIB"
 337..>1917
 /codon_start=1
 /product="Unknown (protein for IMAGE:6879542)"
 /protein_id="AAH54296.1"
 /db_xref="GI:32450291"
 /translation="MSGTQATQTSRTGEARYPPHAYPMQLSRAHTDANLEYOH
 LTBDSYSHLGHSVLLTQRRPSSLTQFQGNRSQELHGRDQHSYHLHMSQKHDL
 HBAKPRLDLHPLMRSLITQNOHAIGEELNKRDMTGLKLEPSPSPPLADGDL
 DLLPRLSKLEELIMDRVDREIIMVEQQITLKKKQOQLEEAARLEPEKPSPP
 VESKHSVLQIIVDENRKAQAARILLEGIPHVLPYNQPSDTKYOHENIKINQM
 RKLLILYFRNRHARKQWEQKLCQRYDQLEANEWEKKVERIENPRRAKSKIREYE
 KQPEIRKQRELOERQMSRVQGNLSAARSEHEVSEIIDLGSQENLERMQRL
 AVPPMLFAEQORIKFINNGLMDPMKYKROVMNMSQEKDTRFKFIQPKN
 FVFIASFLERKTVSDCVLFYLLTKQENYKSLIRNRYRRRGKQNMQPMRTNQEKDEKE
 KEVDKEYKQKKKKKKK"

ORIGIN

Alignment Scores:
 Pred. No.: 3.14e-36 Length: 1917
 Score: 2155.00 Matches: 414
 Percent Similarity: 86.58% Conservative: 57
 Best Local Similarity: 76.10% Mismatches: 55
 Query Match: 16.31% Indels: 18
 DB: 5 Gaps: 3

US-09-522-753-5 (1-2517) x BC054296 (1-1917)

Qy 1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyPro 20
 Db 337 ATGCTGGATCTACTCAGCAAGCCACACAGCGGAGCAGGAGCGGCGGTACCT 396
 Qy 21 ProHisSerLeuSerTyProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
 Db 397 CCCATGCTATGGCTTATCCATGTCAGTGTCTCTGTCGCGCACACGGATGCCA 456
 Qy 41 GluTyThrGlnHisSerArgAspTyAlaSerHisLeuSerProGlySerIleIleGln 60
 Db 457 GAGTACCAGCACACACCCCGGATTAACAGCTCCACCTAGGCGCACGGCTCTGTT 516
 Qy 61 ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln 80
 Db 517 ACAACAAGAAAGAGACCCCTCAGTCTATCTGAATTCAGCCAGGAATGAAGGTCT 576
 Qy 81 GluLeuHisLeuArgProGluSerHisSerTyThrLeuProGluLeuGlyLysSer 99
 Db 577 GAGTTACATGGAGAGAGACCAACATTCCTACTCATCATGACATGACCAACATGAT 636
 Qy 100 MetGluPheIleGluSerTyArgProArgLeuGluLeuLeuProAspProLeuLeuArg 119
 Db 637 CTAGACTTGGCAGAGGCAAGAGCGCTAGATTGGATTCTTCATGACCCACTTATGCGA 696
 Qy 120 ProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArg 139
 Db 697 CACTCA---CTTATAACGCAAAACCAACGATGCAATAGGGGAAGAAATTAACAAGAT 753
 Qy 140 SerLeuThrGlyLysLeuGluProValSerProProSerProHisThrAspProGlu 159
 Db 754 TGCATGACTGGGAAGCTAGAGCCCTGTATCCCTCCAGTCTCTCTCTAGCTGATGGAGAT 813
 Qy 160 LeuGluLeuValProProArgLeuSerLysGluLeuIleGlnAnMetAspArgVal 179
 Db 814 CTGGATCTTTTGGCTACCAGGCTTTTCCAGGAAGAGCTGATCAGAAATATGGACAGATA 873
 Qy 180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGln 199
 Db 874 GATCGGAATATCATCTGTTAGAGCAACAGATCACCAAGCTGAAGAAAAAGACAAACAG 933

```
RESULT 23
AB093281
LOCUS          6328 bp      mRNA      linear      ROD 07-FEB-2003
DEFINITION    Mus musculus mRNA for mKIAA1047 protein.
ACCESSION     AB093281
VERSION       AB093281.1  GI:26006244
KEYWORDS      FLI CDNA.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Hara,Y., Nagase,T.,
Ohara,O. and Koga,H.
Prediction of the coding sequences of mouse homologues of KIAA
gene: 1. The complete nucleotide sequences of 100 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences
of cDNA clones randomly sampled from size-fractionated libraries
Unpublished
2 (bases 1 to 6328)
Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
Direct Submission
Submitted (04-OCT-2002) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics, 2-6-7
Kazusa-kamatori, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.
FEATURES
Location/Qualifiers
1..6328
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbg00559"
/tissue_type="brain"
/dev_stage="adult"
/notes="vector:modified pBC SK+"
1..6328
/gene="mKIAA1047"
<1..5175
/gene="mKIAA1047"
/notes="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence=not_experimental
/product="mKIAA1047 protein"
/protein_id="BAC41465.1"
/db_xref="GI:26006245"
/translation="KASRKPRERDVSCESVASTVSAQEDIEDIASNEENPEDESEG
AENSSTESAPSPVVEAAKSSSEDSSENAASRGNTPEVALEATTDPACASPSSAVP
TTKPAERSEVAQVTDASAETAEPMDVDHERCEAGSVLDPDPATPKADSDVPEMQV
PENTAKSGDKAKERDLESTSEKTEARDEDVVAEGIERPEQSDSDSATCSADSGV
DEPERQORVPMADKPSLLTPPGSILISSIPKPNLULLQLQRAAVIPMWVSQPCFN
IPGTQVSGVALYQRHKAMHESALLEQOEQVDECRSTSCSTKSGKPNRWE
VLQAPHOVITLPEGRVLTPTPTPPPLIPSSKTTTVAPEKPSIMGSKSQGTFG
TVLSSHNOAYPOBAKPKPSVGSISLGLPQOESTKAAPLYIKQEEFSPSSQNSQPEGL
LVRAQHEGVRTAGAVOEGSITRGTPASKIVETISLRSSTTQCTPALPOAGIETE
ALKVGPVSNMPIESPEKVRREAAKSHVIEGKSGHILSYDLSLCRALPRGSPHD
LKERTVLSIMQGTPTAESEFEDGLYKQIKRESPIRAPEGAITKGVYDGIIT
IKEMGRSIHEIPRODILTQESRKTPEVQSTRIPEGSISQGTPIKFDNNSSGSAIKH
NKVSLITGPSKLPRGMLEIVPENIKVVERGYEDVKAPEVRAHRTSVVSSGSEVLRS
TLHEAPKALSPGLYDSDSARPTVSYQNTISRGSPMMNRTSDVSSKASHERKSTL
TTPQRESIPAKSPVPCVDVSVSHSPDPHRSAAACEVVRSHLPDLDPAMPFRHALD
PAAAILLQRLQSLPTPGYSQQLYAMENTROTILNDYITISQOMYNLAPDVTGRLSP
REQLGLPYPATRGIIIDLNTMPTIILVPHAGGSTPEPMRITFTICTQVTFPPPYNA
ASLSPGHPTLAAAASAEERERERERERERERERERERERERERERERERERERER
QPRGSHGVYRSPSVRTOETILQORPSVFGQTNGTIVITLPTAQLRIMPLPSG
GPSISQGLPASRVNTAADALAAALVDAAPQMDVSKTKSEKHEAARLEENLRSGAA
VSEQQOLEOKNLEVERKSVQCCTSSALPSGKAQPHASVVYSEAGDKGPKPSRYEE
ELTRGKTITITANFIDVILITQIASDKDARERGSSSDSSSLSHRYETASDAIEV
ISPASSPAPPQKFPQAIQFDMVKANQENESTFQYEGPLHHYRSQSESPPQQPPLP
AB093281
LOCUS          6328 bp      mRNA      linear      ROD 07-FEB-2003
DEFINITION    Mus musculus mRNA for mKIAA1047 protein.
ACCESSION     AB093281
VERSION       AB093281.1  GI:26006244
KEYWORDS      FLI CDNA.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Hara,Y., Nagase,T.,
Ohara,O. and Koga,H.
Prediction of the coding sequences of mouse homologues of KIAA
gene: 1. The complete nucleotide sequences of 100 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences
of cDNA clones randomly sampled from size-fractionated libraries
Unpublished
2 (bases 1 to 6328)
Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
Direct Submission
Submitted (04-OCT-2002) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics, 2-6-7
Kazusa-kamatori, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.
FEATURES
Location/Qualifiers
1..6328
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbg00559"
/tissue_type="brain"
/dev_stage="adult"
/notes="vector:modified pBC SK+"
1..6328
/gene="mKIAA1047"
<1..5175
/gene="mKIAA1047"
/notes="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence=not_experimental
/product="mKIAA1047 protein"
/protein_id="BAC41465.1"
/db_xref="GI:26006245"
/translation="KASRKPRERDVSCESVASTVSAQEDIEDIASNEENPEDESEG
AENSSTESAPSPVVEAAKSSSEDSSENAASRGNTPEVALEATTDPACASPSSAVP
TTKPAERSEVAQVTDASAETAEPMDVDHERCEAGSVLDPDPATPKADSDVPEMQV
PENTAKSGDKAKERDLESTSEKTEARDEDVVAEGIERPEQSDSDSATCSADSGV
DEPERQORVPMADKPSLLTPPGSILISSIPKPNLULLQLQRAAVIPMWVSQPCFN
IPGTQVSGVALYQRHKAMHESALLEQOEQVDECRSTSCSTKSGKPNRWE
VLQAPHOVITLPEGRVLTPTPTPPPLIPSSKTTTVAPEKPSIMGSKSQGTFG
TVLSSHNOAYPOBAKPKPSVGSISLGLPQOESTKAAPLYIKQEEFSPSSQNSQPEGL
LVRAQHEGVRTAGAVOEGSITRGTPASKIVETISLRSSTTQCTPALPOAGIETE
ALKVGPVSNMPIESPEKVRREAAKSHVIEGKSGHILSYDLSLCRALPRGSPHD
LKERTVLSIMQGTPTAESEFEDGLYKQIKRESPIRAPEGAITKGVYDGIIT
IKEMGRSIHEIPRODILTQESRKTPEVQSTRIPEGSISQGTPIKFDNNSSGSAIKH
NKVSLITGPSKLPRGMLEIVPENIKVVERGYEDVKAPEVRAHRTSVVSSGSEVLRS
TLHEAPKALSPGLYDSDSARPTVSYQNTISRGSPMMNRTSDVSSKASHERKSTL
TTPQRESIPAKSPVPCVDVSVSHSPDPHRSAAACEVVRSHLPDLDPAMPFRHALD
PAAAILLQRLQSLPTPGYSQQLYAMENTROTILNDYITISQOMYNLAPDVTGRLSP
REQLGLPYPATRGIIIDLNTMPTIILVPHAGGSTPEPMRITFTICTQVTFPPPYNA
ASLSPGHPTLAAAASAEERERERERERERERERERERERERERERERERERERER
QPRGSHGVYRSPSVRTOETILQORPSVFGQTNGTIVITLPTAQLRIMPLPSG
GPSISQGLPASRVNTAADALAAALVDAAPQMDVSKTKSEKHEAARLEENLRSGAA
VSEQQOLEOKNLEVERKSVQCCTSSALPSGKAQPHASVVYSEAGDKGPKPSRYEE
ELTRGKTITITANFIDVILITQIASDKDARERGSSSDSSSLSHRYETASDAIEV
ISPASSPAPPQKFPQAIQFDMVKANQENESTFQYEGPLHHYRSQSESPPQQPPLP
```

PSSQSEGMGVPRTHRLITLADHICQIIITQDFARNQVPSQSTSTFOTSPSALSSTPV
 RTKTSYSRSESQSVLHPRGPRVSPENLVKDSRGRCKSPKSHIPSEBVEPTS
 PPOGPAVHERKQSDSMLLSQSDVPAEORSRSPGISYLPSPFTKLESTSPMVKSKK
 OEIPFRKLNSGGGSDMAAOPGTEIENLPAVTSAGVSRKSHSFADPASNGLGEDI
 RKALMGSDDKVEDHGVHPGIMPGSASTSVVTSSEARRDEGESPHAGVCKPKL
 INKSNRSKSPIPGQSYLGTERTSPSSVSHSEGDYHROTQPGHAWERDRPSSTGTPP
 INPLTIRMLSSPTPTQIACAPSAITQAAPHQNKRIWEREPAPLLSAQYETLSDD

ORIGIN

Alignment Scores:
 Pred. No.: 1,58e-34 Length: 6328
 Score: 2091.50 Matches: 671
 Percent Similarity: 46.13% Conservative: 253
 Best Local Similarity: 33.50% Mismatches: 638
 Query Match: 15.83% Indels: 441
 DB: 10 Gaps: 92

US-09-522-753-5 (1-2517) x AB093281 (1-6328)

Qy 672 LysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAlaAlaSer 691
 Db 13 AAACCCCGTGAGGAACAGAGATGATCTCAGTGTGAAGTGTGCTTCCACTGTTCTGCC 72
 Qy 692 GluGluAlaAlaPheProProValValGluAspGluMetGluAlaSerGlyValSer 711
 Db 73 CAA-----GAGGATGAAGATATTGAAGCCTCA----- 99
 Qy 712 GlyAsnGluGluMetValGluAlaGluAlaLeuHisAlaSerGlyAsnGluVal 731
 Db 100 ---AATGAGAGGAAATCAGAGATAGTGAAGT----- 132
 Qy 732 ProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIle 751
 Db 133 -----GCTGAAATAGTTCTGATACAGAAAGTGCT 162
 Qy 752 ProSerProHis---ThrGluAlaAlaLys-----AspThrGlyGlnAsnGlyPro 767
 Db 163 CCCTCTCCTTCCACAGTGAAGTGCACCAAGTCCAGTGAAGACACAGCAGTGAATAATGCT--- 219
 Qy 768 LysProProAlaThrLeuGlyAlaAspGlyProProGlyProProGlyProProAlaArg 787
 Db 219 ----- 219
 Qy 788 ArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThr 807
 Db 220 ---GCTTCTCGAGGAACACCGACCTGTGCTGAGCTTGAGGCCACCACTGACCTGCA 276
 Qy 808 ProProProAlaProProSerProSerAlaProProValValProLysGluGluLys 827
 Db 277 CCCTGTGCATCTCC-----TCTTCAGCAGTTCCAAACCACAAACCAGCAGAAAGGAA 330
 Qy 828 GluGluGlu-----ThrAlaAlaAlaProProValGluGlu 839
 Db 331 AGCGTGAGGCCACAGGTGACGACGCGCCAGTCTGAGACCCGAGCCGATGGAGCTA 390
 Qy 840 GlyGluGluGln-----LysProProAlaAlaGluGlu 850
 Db 391 GACCATGAGGAGTGCAGTGCAGGCGCAGTCTGTTCTTGTATCCACCCAGCCCTACCAA 450
 Qy 851 LeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAla 870
 Db 451 -----GCCGACTCCGTGGACCCAGAAATGCAGTA---CCAGAAATACTGGCTCTAAA 501
 Qy 871 GluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAla 890
 Db 502 GGTGAAGGGATCCCAAGGAAGAGACTTGGAGACAC----- 540
 Qy 891 LeuLysAlaGluLysGlyGlySerGlyArgAlaThrThrAlaLysSerSerGly 910
 Db 541 -----AGTGAGAAGACAGAGCTAGAGATGAAGACCTGGTGTGCTGAACAGATCGAG 594
 Qy 911 AlaProGln-----AspSerAspSerSerAlaThrCysSerAlaAspGluValAsp 927
 Db 927 ----- 927

Db 595 AGGCCTGAGCCACAGTCAGTCAGACGACCTCCAGTGCCACTTGCAGTGGCGATGAGGGGTG 654
 Qy 928 GluAlaGluGlyGlyAspLysAsnArgLeu-----LeuSerProArgProSerLeuLeu 945
 Db 655 GATGAGAG---CCAGAGAGGCGAGAGTGTTCCTCCATGGATGCAAGACCTTCATTGTTA 711
 Qy 946 ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeuAspLeu 963
 Db 712 ACTCTCTCGATCTATCTCTAATC---TCATCCCTATTAAACCAACCCCATTCGATCTG 768
 Qy 964 LysGlnLeuLysGlnArgAlaAlaAlaProProIle----- 976
 Db 769 CCACAGCTTCAGCATCGAGCTGCTGTTATTTCACCAATGGTTCCTGCATCCCATGTAAT 828
 Qy 977 -----GlnValThrLysVal 981
 Db 829 ATACCAATTGGAACGCCCTAAGTGGCTATGCTCTTTACCAACGCGCATTTAAGGCCATG 888
 Qy 982 HisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro 1001
 Db 889 CATGAGTCAGCACTCTCTGGAG-----GAG 912
 Qy 1002 ProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGly 1021
 Db 913 CAGCGGCGAGGCAAGAACAGGTAGACTTGGAAATGCAGAACTCTACAGCCCATCAGC 972
 Qy 1022 LysSerArgSerProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLys 1041
 Db 973 ACTTCTAAGAGTCCA-----AACAGGAG----- 996
 Qy 1042 LeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProProArgGlu 1061
 Db 997 -----TGG-----GAA 1002
 Qy 1062 ValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTrpAlaProProGly 1081
 Db 1003 GTCTCTCAACCTGCTCGCAT-----CAAGTGATAACTTAACCTTCTCTGAAGGG 1050
 Qy 1082 HisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThr 1101
 Db 1051 GTTCGGGTCCA-----ACAACGAGCAACACGAGCCACCACTTCCCTC 1095
 Qy 1102 IleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGln 1121
 Db 1096 ATCCCATCATCTAAACCCACAGTGGCATCAGAAAA---CCATCTTTTATA-----ATG 1146
 Qy 1122 IleGlyAlaIleSerGlnGly-----MetSerValGlnLeuHisValPro 1136
 Db 1147 GAGGGGTCTATCTCACAGGGAACCTCTGCGCACTTACTTGTCTTCTCATATAATCAGGCTTAT 1206
 Qy 1137 TyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMet 1156
 Db 1207 CCACAGAACCCCTAAGCCCTCGTGGGTCTATCTCTGGATTGCCCGGCAGCAG 1266
 Qy 1157 AspProLysLysLeuAlaProPheSerGlyLysGlnGlnLeuSerProArgGly 1176
 Db 1267 GAGTCTACCAAAACAGCTCTCTTTGACCTACATCAAGCAGAGAAATTTTCTCCGAGAAGC 1326
 Qy 1177 GluAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArg 1196
 Db 1327 CAAAACCTCAACCTTCTGAGGTTTATTGGTC---AGAGCGCAGCATGAAGGTGTGGTCAGA 1383
 Qy 1197 GlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArg 1216
 Db 1384 GGCACCTGCA---GGGCTGTCCAAGGAAGGAAGTATACTCGGGAACCTCCAGCCAGCAAA 1440
 Qy 1217 ValProSerAspSerAlaIleThrThrArgGlySerIleThrHisGlyThrProAla--- 1235
 Db 1441 ATCTCAGTGGAGACCATTTTCATCGTTCGGGGCTCTATTACCCAGGGGACCCCGAGCTGT 1500
 Qy 1236 -----AspValLeuTyrLysGlyThrIleThrArgIle---Ile 1247
 Db 1501 CCCAGGCTGGAATACCAACAGAGGCTTTGGTGAAGGAGCCTGTCTCCAGGATGCCTATT 1560

Qy	1248	GlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHis	1267
Db	1561	GAAGAAAGCAGTCTCTGAGAGGTC-----AGAGAGGAAGCTCATCCAAGGCCAT	1611
Qy	1268	ValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThr	1287
Db	1612	GTTATCTATGAAGCAAAAGTGGACATATCTTTATCATAT-----	1650
Qy	1288	GlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaPro	1307
Db	1650	-----	1650
Qy	1308	LysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIle	1327
Db	1650	-----	1650
Qy	1328	GluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHis---HisLeuLys	1346
Db	1651	GATAGTCTGATGCCAGGACATTACCC-----AGGGGAGCCCTCATTTGTGACCTCAAA	1704
Qy	1347	GluGlnHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAla	1366
Db	1705	GAGAGGACTGTGCTGTCTGCTTCCATAATCGAGGCACACCAAGAGCCAGCGGAAGC	1764
Qy	1367	GlnGluAspTyrIleuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProPro	1386
Db	1765	TTTGAGGACGCGCCTT---AAATACCCCAACAGATATAAAGGAGAGC-----	1809
Qy	1387	ProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLys	1406
Db	1810	---CCTCCCATCCGAGCATTTGAAGTGCCATT-----ACCAA	1845
Qy	1407	LeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHis	1426
Db	1846	GGAAAAACCA---TACGATGCT---ATCACCAACCATCAAAGAAATGGGGCGCTCCATCCAT	1899
Qy	1427	GluIleProArgGluGluLeu-----ArgHisThrProGluLeuProLeu	1441
Db	1900	GAATCCACCGGCAAGATATTCTTAATCTCAGAAAGCCGGAATACTCAGAAGTGGTCCAG	1959
Qy	1442	AlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThr	1461
Db	1960	AGCAGAGGCCAATAATTGAGGTTCCATTCCAGGGCCACACCAATAAATTTGACAAC	2019
Qy	1462	GlyAlaSerThrThrGlySerLysLysHisAspValArgSerIleLleGlySerProGly	1481
Db	2020	AAC---TCAGGTCAATCAGCTATCAAAACAAATGTGAAGTCCTTAATCAACAGCGCGCTAGC	2076
Qy	1482	ArgThrPheProProValHisProLeuAspValMetAlaAsp---AlaArgAlaLeuGlu	1500
Db	2077	ANACTA-----CCCGTGGNATCTCGAATAATTGTACACAGAACAATAAAGTAGTAGAA	2130
Qy	1501	ArgAlaCysTyrGlu-----GluSerLeuLysSerArgProGlyThrAla	1515
Db	2131	CGGGCAAAATATGAGGATGTGAAAGCAGGCGAGCCAGTGGCGAGCCCGGCACAGCTCAGTG	2190
Qy	1516	SerSerSerGlyLysIleAlaArgGlyAlaProValIleValProGluLeuGlyLys	1535
Db	2191	GTGAGCTCTGGCCCTCCGTTCTCAGGTCTACA-----CTTCAGAGACTCCCAAA	2241
Qy	1536	ProArgGlnSerProLeuThrTyrGluAspHisGlyAla-----ProPhe	1550
Db	2242	GCACAGCTGAGCCCGGACTCTATGATGACAGCAGCGCTCGCAGGACTCCTGTGAGCTAC	2301
Qy	1551	AlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGln	1570
Db	2302	CAGAACCACTATCCAGAGGCTCCCATGATGAACAGA-----	2340
Qy	1571	GluGlySerLeuSerSerSerLys---AlaSerGlnAspArgLysLeuThrSerThrPro	1589
Db	2341	ACTTCTGATGTTTCTTCAGCAAGTCTGCCAGTCAATGAAGGAATCAACTCTGACCCCA	2400

Qy	1590	ArgGlu-----llealalysSerProHisSerThrValProGluHisHisPro	1600
Db	2401	ACCCAAAGAAAGTATACACGCAAGTCTCCAGTGCCTCCGGGGTGGATCCCGTCGTGAGT	2460
Qy	1606	HisProIleSerProTyrGluHisIleuLeuArgGlyValSerGlyValAspLeuTyrArg	1625
Db	2461	CAC-----AGCCCAVTTGATCTCTCAACAGAGTAGCGCTGCAGGAGAGGTTTATCGG	2514
Qy	1626	SerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAsp---	1644
Db	2515	AGCACACTACCCAGCAGCTTGGATCCA---GCTATGCCCTTTCACAGGCGTTTGGATCCT	2571
Qy	1645	AlaAlaAlaIleTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeu	1664
Db	2572	GCAGCTGCTGCTTACCTGTTTACAGACAGACAGCTTTCACCAACCCAGGATACCCAGATCAG	2631
Qy	1665	TyrProProTyrIleuIleargGlyTyrProAspThrAlaAlaIleuGluAsn---ArgGln	1683
Db	2632	TACCAGCTCTAT-----GCAATGGAGAATACAGAGCAG	2666
Qy	1684	ThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMet	1703
Db	2665	ACAATCTCTCAACATTACATTACTCTACAGCAGATCGAGTGNATCTG-----	2712
Qy	1704	AlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgIleuSerSerIleuAlaLeuAsn	1723
Db	2713	-----CGCCCTCATGTCTACACGGGACTGTCCCCACGAGCAGCAGCCTATGGGGCTCCCT	2766
Qy	1724	TyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisIleuProValLeu	1743
Db	2767	TACCCAGCT---ACAAGAGGAATCAVTTGACCTTGACCAATATGCCTCCA---ACAACTCTTA	2820
Qy	1744	ValProProThrProGlyThrProAlaThrAlaMetAspArgIleuAlaTyrLeuProThr	1763
Db	2821	GTGCCTCATCGAGGGGAACGACACCCCTCCCATCGACAGATCATGTTATCTCTGTT	2880
Qy	1764	AlaProGlnProPheSerSerArg---HisSerSerSerProIleuSerProGlyGlyPro	1782
Db	2881	ACACAGTTACTTTCCCTCCAGGCCATATAACGCTGCTTCTGTCTCCAGGACACCCA	2940
Qy	1783	ThrHisIleuThrLysProThrThrThrThrSerSerSerGluArgGluArgAspArg	1802
Db	2941	ACACACCTT-----GCAGCAGCTGCAAGTGTCTGAGAGGGAACGAGAGCGGAAAGG	2991
Qy	1803	GluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrValGlu	1822
Db	2992	GAGAAGGAGCGCAACGTGTAGCGCGAGCGTGAACGCTGAACCGGAAAGATCGCT	3051
Qy	1823	HisAlaPro-----IleTyrArgProGlyThrGluGlnSerSerGlySerSerGly	1839
Db	3052	GCTGCTCCCGCTGACCTTACCTACGACCAAGGTTCAAGACAG-----	3093
Qy	1840	SerSerGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln	1859
Db	3094	-----CCAGGCGCTCTGGCAGCCACCGGATATGTCGCTCC	3129
Qy	1860	HisSerProIleSerProArgThrGlnAspAla---LeuGlnGlnArgProSerValLeu	1878
Db	3130	CTTTCCCT---TCAGTAAGAATCTCAGAGACCATCTCTGCAACAGAGACCCAGTGTTTC	3186
Qy	1879	HisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu	1898
Db	3187	CAGGGCACCAATGGAACCAAGTGTAAATCACACCTTTGGACCCAACTGTCTACGCTACGCATC	3246
Qy	1899	ArgSerThrSerThrSerProVal-----ArgProAlaAlaThrPhePro	1914
Db	3247	ATGCCACTGCTCTCTGGGGGCCCTTCCATAAGTCAAGCGCTGCAGCGCTCCCGTTATCAAC	3306
Qy	1915	ProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGlu	1934
Db	3307	ACTGCTCGGATGCC---CTGGCTGCTCTTGTGGATGCTCAGCTTCTGCACCCAGATG	3363
Qy	1935	ProValLeuLeuProLysGluAlaProArg-----	1944

Db 3364 GATGTTTCCAAAACAAAGACAGATGAAGCATGAGCTGCCAGGTTAGAGAAAAATTTGAGA 3423
 Qy 1945 -----ValAlaArgProGluArgProAlaAspThrGlyHisAlaPheLeuAla 1961
 Db 3424 AGCAGGTGACAGCAGCTTAGTGAACAGCAGCAGCTAGACGAGAAACCTGGAGGTGAG 3483
 Qy 1962 LysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluPro 1981
 Db 3484 AAGAGATCTGTTCAAGTGTGTTGCACTTCTTCAGCCCTTCCCAAGTGGCAAGGCCAGCCT 3543
 Qy 1982 ArgProLeuVal-----ProProValSerGly 1990
 Db 3544 CATGCCCTCAGTAGTGTATTCTGAGCGTGGAAAGATAAAGGCCCTCTCTCCAAAATCCAGA 3603
 Qy 1991 HisAlaThrIleAlaArgThrProAlaLysAsn-----LeuAlaProHisHisAlaSer 2008
 Db 3604 TATCAGGAGAGCTAAGACCCGAGGAGGAGACTACCATCTACTGACGTAACTTCATAGAC 3663
 Qy 2009 ProAspProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSer 2028
 Db 3664 GTGATCATCACCCGCGAAATTCCTCGACAAAGATCGAGGGAACGTGGCTCTCAAGT 3723
 Qy 2029 LysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyr 2048
 Db 3724 TCAGACTCTTCT-----AGTAGCTTGTCTCTCACAGGTATGAAACG 3765
 Qy 2049 SerProGluGlyValGluProValSerProValSerProSerLeuThrHisAspLys 2068
 Db 3766 GCTAGTGCCATTGAGTGATGAAGTCCCGCCAGCTCACCTCGCACCCAGGAAAG 3825
 Qy 2069 GlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgPro 2088
 Db 3826 CCACAGCCCTATCAGCCAGACATGTTAAGCAAAATCAAGCAGAAATAGTCCACTCGA 3885
 Qy 2089 LysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArg--- 2107
 Db 3886 CAGTATGAAGGTCCA-----CTGCATCATTTATCGGTCC 3918
 Qy 2108 -----ProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThr--- 2122
 Db 3919 CAGCAGGAATCACCATTCTCACACAAACAGCCACCTCGCCCTTCTCCAGTCAGAG 3978
 Qy 2123 ---AlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGlu 2141
 Db 3979 GGAATGGACAGGTGCCAGGACCCATCGACTGATCACACTTGTGTGACCATCTGTCAA 4038
 Qy 2142 ValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAla 2161
 Db 4039 ATTATCACACAAGATTTTGTAGAAAT-----CAAGTTCCTCGCAGCCTTCTACTTCT 4092
 Qy 2162 ProLeuTyrSerPheProGlyAla-----SerCysProValLeuAspLeuArgArgPro 2179
 Db 4093 ACATTCCAAACTTACCATTCTGCTTGTATCCACACCTGTG-----AGAACTAAA 4143
 Qy 2180 ProSerAspLeuTyrLeuProProAsp-----HisGlyAlaProAla 2194
 Db 4144 ACCTCAGCCGCTACACCCAGATCAGTCTCAGACTGCTGTGCATCCAGACCAAGT 4203
 Qy 2195 ---ArgGlySerProHis-----SerGluGlyLysArgSerProGlu 2208
 Db 4204 CCTAGAGTCTCTCCAGAAATCTTGTGGATAAATCCCGGAGAGCAGCGCTGGAAATCT 4263
 Qy 2209 ProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProPro 2228
 Db 4264 CCAGAGGAGTATATC-----CCATCAGAGCCCTATGAGCCCATCTCCCCACC 4314
 Qy 2229 GluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAsp 2248
 Db 4315 CAGGCCCTCTGTG-----CATCAGACGACGACGATGTGCTCTTGTACACAGAG 4368
 Qy 2249 GlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnPro 2268

Db 4369 GGAGTGGACCTGCTGAGCAAAAGGAGTGATTCTTCGATCACCCAGGAAGTATTAAGCTACTTG 4428
 Qy 2269 ProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGln 2288
 Db 4429 CCTTCATCTTCCACCAAGCTT---GAAAGGCATCATCCCATGGTTTAATCAAAGAAACAG 4485
 Qy 2289 GluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGln 2308
 Db 4486 GAAATTTTCTGTAAGTTGAACCTCTCTCTGGTGGAGGTGACTCTGATATGGCAGCTGCTCAG 4545
 Qy 2309 ProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArg 2328
 Db 4546 CCAGGAACACAGATCTTCAATCTGCCAGCAGTTACCACATCAGGTGAGTCCAGTCAAGA 4605
 Qy 2329 SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLys 2348
 Db 4606 AGCCATCTCTTGTGCTGATCCGCCAGT---AACCTTGGTCTAGAAGACATCATCAGAAAG 4662
 Qy 2349 AlaLeuMetGlyLysTyrAspGlnTrpGluGlu-----SerProPro 2362
 Db 4663 GCTCTCATGGGAAGTTTGTATGATAAAGTTGAAGATCATGGTGTGTGTCATGTCCTCCATCCT 4722
 Qy 2363 LeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPro 2382
 Db 4723 GTG-----GSCATTATGCTGCTGTAGTCCAGCACCTCAGTG----- 4758
 Qy 2383 IleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGly---Gly 2401
 Db 4759 GTGACGAGCAGCAGGCGCAGAGATGAAGGGAGGCATCACCCTCATGCGAGGTATGC 4818
 Qy 2402 LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeu 2421
 Db 4819 AAACCAAGCTGATCAACAAATCAACAGCAGGAAGTCTAAATCTCTTATTCCTGGGCAA 4878
 Qy 2422 Ala-----SerGlyAspArgProProSerValSerSerValHisSerGluGlyAspCys 2439
 Db 4879 AGCTATTTAGAACTGAAAGGCCCTTCTTCTCTCTCTCTGTCATTCAGAAAGGTGATTC 4938
 Qy 2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
 Db 4939 CACAGGCGACACCA-----GGATGGGATGGGAAGATCGGCCCTTCTCAACAGGTCT 4992
 Qy 2460 ThrProPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro 2479
 Db 4993 ACTCAGTTCTCTTACACCTCTGACCATACGGATG-----CTCAGCAGTACA 5040
 Qy 2480 ProProProGlyLeuProAlaGlySerGlyProLeu-----AlaGlyProHis----- 2495
 Db 5041 CCACCTACACAGATCGCATCGCCCTCTGTCATCACCACCAAGCAGCTCCACATCAACAG 5100
 Qy 2496 ---HisAlaTrpAspGluGluProLysProLeuLeuCysSerGlnTyrGluThrLeuSer 2514
 Db 5101 AACCGCATCTGGAGAGGGAGCGCTGCCCGCTCTCTTCAGCGCATGATGAGACTGTCT 5160
 Qy 2515 AspSerGlu 2517
 Db 5161 GATAGTGAC 5169
 RESULT 24
 AB019524
 LOCUS
 DEFINITION Homo sapiens mRNA for nuclear receptor co-repressor, partial cds.
 ACCESSION
 AB019524
 VERSION
 AB019524.1 GI:4519939
 KEYWORDS
 nuclear receptor co-repressor.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (sites)
 Nagaya,T., Chen,K.S., Fujieda,M., Ohmori,S., Richer,J.K.,
 Horwitz,K.B., Lupski,J.R. and Seo,H.
 Localization of the human nuclear receptor corepressor (hN-CoR)

gene between the CMT1A and the SMS critical regions of chromosome

17p11.2
Genomics 59 (3), 339-341 (1999)
MEDLINE
99375328
1044336
PUBMED
2 (bases 1 to 6541)
REFERENCE
Nagaya, T.
Direct Submission
Submitted (04-NOV-1998) Takashi Nagaya, Research Institute of
Environmental Medicine, Nagoya University, Department of
Endocrinology and Metabolism, Furo-cho, Chikusa-ku, Nagoya, Aichi
464-8601, Japan (E-mail: tnagaya@iem.nagoya-u.ac.jp).
Tel: 81-52-789-3867, Fax: 81-52-789-3887)

FEATURES

source

1. 6541
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="chromosome 17"
/map="17p11.2"
/cell_line="Hela cell"
<1..4404
/codon_start=1
/product="nuclear receptor co-repressor"
/protein_id="BA075814.1"
/db_xref="GI:4519940"

CDS

/translation="ESQRQEQDLECRSSTSPGCTSKSPNREWEVLQAPHOVITN
LPEGVRLPTTRPPPLIPSSKTTIVASEKPSFIMGSSISQGTPTGTYLITSHNQASYT
QETPKPSVGSISGLRQQRSAKSATLPIYKQEEFSPRSQNSQPEGLVRAQHEGVVP
GTAAIGEGSITRTGTSKISVESIPSLRGSITQGTTPALPQTGIPTEALVKGSISSRP
LEDSNPKGHEEASGHVIEGSGHLSYDNIKNAREGTRSPRTAHEISLKRYSYES
VEGNKCKMSRSPVSAPEGLICRALPGSPHSDLKERTVLSGSIIMQGTTPRATES
FEDGLPKQIKRESPIRAPEGAITKGYDGIITIKEMGRSIIHEIPRODILLTORSR
KTEPVQSTRIIEGSIQGTPIKFNNSGSAIKNVKSLITGPKSLRGMPEPLIIV
PENIKVVERGYEDVKAQETVRSHTSVSSGSPVLRSLHEAPKAQLSPGIYDDTSA
RRTPVSQNTMGKSDPMNRTSDVTISNNKSTNHERKSTLTPQRESIPAKSPVPGVD
PVVSHSPFDHRRSTAGEVVRHLPLHDPAMPFHRALDPAAAYLFORGLSPTPGY
PSQOYLAMENTQTLINDVITSQOMOVNLRPDVARGLSPREOPLGLPYPATRGIIIDL
TNMPTLIVHPHGTSITPPMDRTIYIFGTQITPPRPYNASMSGPHPTHAAASAE
REREREKREKRIIAASDRLYLRGSEQGRPGSGHGVRSFSPSVRTQETWMLQOR
PSVPGNGTSVITPLDPTAOLIMPLPAGGSPISQGLPASRINTADALALVDAAA
SAPQMDVSKTKSEHEARLEENLRSAVSEQQLEOKLTLEVRKRSVQCLYTSAP
PSGKPPSHSVVSEACKDGPKKRYEELRGTKTITTAANFDIVITROIASDK
DAREGQSDSSSLSSHRVETPSDAIEVISPASSAPPOEKLQTVQPEVWANOAE
NDPROEGPLHVRPOQSPSPQOOLPPSQAEGMGQVPRTHRLITLADHICQITQ
DFARNQVSSQTPQPPSTIFQNSFSAVSPVTKSNRYSPSSQAQSVHQRGSRV
SPENLVKRSRGPSPKSPERSHVSSEPIEPIPPQVPVHKKQDLSLLLSQRAEPAE
QRNDARSPTGSITPLPFTKLENTSPMVKSKQIEIFKLNSGGGSDMAAQPGETI
PNLPAVTTSGSVSRGHSFADPNLGLIEDIIRKALMGSDFDKVEDHGVVMSQPMGVV
PGTANTSVTSGETRRREGDPSHSGGVCKPKLISKNSRKSPIPGQGYLGTFRPS
SVSVSHSEGDYHRTQPGWMDRPSSTGSGTQFPYNPLTMRMLSTPTPTIACAPSAVN
QAAPHQNRIRWEREPAPLLSAQYETLSDSD"

polya_site

/note="47 A nucleotides"

ORIGIN

Alignment Scores:
Pred. No.: 1,21e-30 Length: 6541
Score: 1900.00 Matches: 581
Percent Similarity: 49.36% Conservat: 226
Best Local Similarity: 35.54% Mismatches: 530
Query Match: 14.38% Indels: 298
DB: 9 Gaps: 79

US-09-522-753-5 (1-2517) x AB019524 (1-6541)

QY 1006 GlnProGluSerAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSer 1025
Db |||||
19 CAAGAACAGATAGATTGGAATGTAGAAAGTCTCAAGTCCATGCGCACATCAAGAGT 78
QY 1026 ProAlaProAlaAspLysGluAlaAlaGluAlaGlnLysLeuProGlyAsp 1045
Db |||||
79 CCA-----AACAGAG----- 90

QY 1046 ProProCysTyrThrSerGlyLeuProPheProValProProArgGluValIleLysAla 1065
Db |||||
91 -----TGG-----GAATGCTTTCAGCCT 108
QY 1066 SerProHisAlaProAspProSerSerAlaPheSerTyrAlaProProGlyHisProLeuPro 1085
Db |||||
109 GCTCCACAT-----CAAGTGATAACTAATCTCCCTGAAGCGCTTCGCTTCG 156
QY 1086 LeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPro 1105
Db |||||
157 -----ACAACTCGACCAACCGGCCACCGCCCTCTCTCATTCGCTCATCC 201
QY 1106 ProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIle 1125
Db |||||
202 AAACACACAGTGGCTTCAGAAAAA---CAATCTTTTATA-----ATGGGAGGCTCCATC 252
QY 1126 SerGlnGly-----MetSerValGlnLeuHisValProTyrSerGluHis 1140
Db |||||
253 TCACAGGGAACACACAGGCACATTATTGACTTCTCTATAATACAGGCTTCTACACTCAAGAA 312
QY 1141 AlaLysAlaPro---ValGlyProValThrMetGlyLeuProLeuProMetAspProLys 1159
Db |||||
313 ACACCCAAAGCCGTCAGTGGGATCTATCTCTTGACATGCCACGCCAACAGCAAAATCTGCC 372
QY 1160 LysLeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGly 1179
Db |||||
373 AAATCAGCTACTTTCCTTACATCAAGCAGGAAGAAATTTCTCCCGAAGCCCAAACTCA 432
QY 1180 ProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAla 1199
Db |||||
433 CAACCTTGAGGGTCTGTGTGTC---AGGGCCCCAACATGAAGGTGTAGTCAGAGGTACCCGA 489
QY 1200 LeuGlySerValProGlyLysSerIleThrLysGlyIleProSerThrArgValProSer 1219
Db |||||
490 ---GGAGCCATCAAGAAGAGTATACTCGGGAACTCCAAACCAGCAAAATTTTCAGTG 546
QY 1220 AspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAla----- 1235
Db |||||
547 GAGAGATTCATCCCTACGCGGCTCTATCACTCAGGGCACCACCGGCTCTGCCCCAGACT 606
QY 1236 -----AspValLeuTyrLysGlyThrIleThrArgIleGlyGluAspSer 1251
Db |||||
607 GCATACCAACAGAGCTTTGTGAAGGGTCCATTCAGAAATGCCCATGTTATTATGAA 720
QY 1252 ProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGlu 1271
Db |||||
667 AGTCCT-----GAGAAAGCAGAGAGAAAGCTGCATCCAAAGGCCATGTTATTATGAA 720
QY 1272 GlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLys 1291
Db |||||
721 GGCAAAAGTGGACATATCTTGTTCATATGATAATATTAAGAAATGCC----- 765
QY 1292 GluAspGlyArgSerSerGlyProProHisGlyThrAlaAlaProLysArgThrTyr 1311
Db |||||
766 CGAAGAGGAGTACAGGACTCAAGAACAGCTCATGAATCAGTTA---AAGAGAGCTAT 822
QY 1312 AspMetMetGlyArgValGlyArgAlaIleSer-----SerAla 1325
Db |||||
823 GAATCAGTGAAGAAATATAAGCAAGGAGTCTCAATGAGGAGTCTCTCTGTATCAGCA 882
QY 1326 SerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHis---His 1344
Db |||||
883 CCGTTAGAGGGCTGATATGCCGAGCATTAACC-----AGGGGGAGTCTCTCATCTTGAC 936
QY 1345 LeuLysGlnGlnHisIleAlaArgGlySerIleThrGlnGlyIleProArgSerTyrVal 1364
Db |||||
937 CTAAAGAAAGAGTCTGTTGCTGCTGCTCAATATGAGGGGACACCAAGCAACAACACT 996
QY 1365 GluAlaGlnGluAspTyrLeuArgGluAlaLysLeuLysLeuArgGlyThrPro 1384
Db |||||
997 GAAAGCTTTGAAGATGGCTTT---AAATATCCCAAAATAATTAAGGGGAAGT----- 1047

Qy 1385 ProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyPro 1404
Db 1048 -----CCTCCATACCAGCAATTTCAAGGTGCCATT----- 1077
Qy 1405 LeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSer 1424
Db 1078 ACCAAGGAAACCA---TATGATGCG---ATCACCAACCATCAAGAAATGGGGCGTTCC 1131
Qy 1425 IleHisGluIleProArgGluGluLeu-----ArgHisThrProGluLeu 1439
Db 1132 ATTCATGAGATTCACAGCGAAGATATTAACTCAGGAAGTCCGAAATCTCCAGAAGTG 1191
Qy 1440 ProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyr 1459
Db 1192 GTCCAGAGCACACGGCGCAATATGAGGTTCATTTCCAGGGCACACCAATAAAGTTT 1251
Qy 1460 AspThrGlyAlaSerThrGlySerLysLysHisAspValArgSerLeuIleGlySer 1479
Db 1252 GACAACAC---TCAGTCAATCTGCCATCAACACCAATGTCAATCTTAAATCACCGGG 1308
Qy 1480 ProGlyArgThrPheProProValHisProLeuAspValMetAlaAsp---AlaArgAla 1498
Db 1309 CCTAGCAAACTATCCCGTGGATGCTCCGCTGGAAATTTGGCCAGAAACATAAAGTG 1368
Qy 1499 LeuGluArgAlaCysTyrGlu-----GluSerLeuLysSerArgProGly 1513
Db 1369 GTAGAACGGGMAAATATAGGATGTGAAGCAGCGGAGACCGTGGCTTCCCGGCACAG 1428
Qy 1514 ThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeu 1533
Db 1429 TCAGTGTAGAGCTGTGCCCTCCGTTCTTAGGTCCACA-----CTGCATGAAGCT 1479
Qy 1534 GlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla----- 1548
Db 1480 CCNAGCACACTGACCCCTGGATTATGATACACACCGAGCCGAGCCCTCTGTG 1539
Qy 1549 ProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrProArg 1568
Db 1540 AGTTATCAAAACACCATGTCCAGAGGCTCACCCATG-----ATGAACAGA 1584
Qy 1569 LeuGlnGlySerLeuSerSerLysAla---SerGlnAspArgLysLeuThrSer 1587
Db 1585 ACTTCTGATGTACAATTTCTTAAACAGTCTACCAATCATGAAAGGNAATCGACACTG 1644
Qy 1588 ThrProArgGlu-----IleAlaLysSerProHisSerThrValProGluHis 1603
Db 1645 ACCCTACCAGGGGAAAGTATCCAGCGAGTCTCCAGTGCCTGGGGTGGACCTGTC 1704
Qy 1604 HisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeu 1623
Db 1705 GTGAGCCAC-----AGTCCGTTTGATCCCATCACAGAGCGACAGCTGCAGGCGAGGTT 1758
Qy 1624 TyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeu 1643
Db 1759 TATCGGAGCCACTGCGCCACGCACTTGATCCA---GCCATGCCCTTTTCAGGGCTTTG 1815
Qy 1644 Asp---AlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrPro 1662
Db 1816 GATCCTGCAGCGGTGCTTACTCTGTTTCAGAGACAGCTTTCACCACTCCAGGTACCCA 1875
Qy 1663 HisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn--- 1681
Db 1876 AGTCAGTATCAGCTTTTAC-----GCAATGGAGAACACA 1908
Qy 1682 ArgGlnThrIleLeuAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThr 1701
Db 1909 AGACAGACAATCTTAAATGATTATCATCTCAACACAGATGCAAGTGAACCTTG----- 1962
Qy 1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla 1721
Db 1963 -----CCTCAGATGTGGCCAGGAGACTCTCCCAAGAGAGACGCACTGGGT 2010
Qy 1722 LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro 1741

Db 2011 CTCCCATACCAGCA---ACGAGAGGAATCATTTGACCTGACCAATATGCTCTCCA---ACA 2064
Qy 1742 ValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeu 1761
Db 2065 ATTTTAGTGTCTCATCCAGGGGAAACAGCACTCTCCCATGGACAGATCTATTAT 2124
Qy 1762 ProThrAlaProGlnProPheSerSerArg---HisSerSerProLeuSerProGly 1780
Db 2125 CCTGGTACACAGATTACTTTCCCTCCAGCGCGGTACAACTCTGCTTCCATGCTCTCCAG 2184
Qy 1781 GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArg 1800
Db 2185 CACCCAAACACACCTT-----GCAGCTGCTGCAAGTGTCTGAGAGGAACCGGAACGG 2235
Qy 1801 AspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThr 1820
Db 2236 GAGCGGAGAGAGGAGCGGAGCGGAGAACGG-----ATTGCTGAGCTTCTCTCCGAC 2286
Qy 1821 ValGluHisAlaProIleThrArgProGlyThrGluGlnSerSerGlySerGlySer 1840
Db 2287 CTC-----TACCTGCGCGCAGGCTCAGAACAG----- 2313
Qy 1841 SerGlyGlyGlyGlySerSerArgProAlaSerHisSerHisAlaHisGlnHis 1860
Db 2314 -----CCTGCGCGACCTGGCAGTCTGATATGTTGCTCCCT 2352
Qy 1861 SerProIleSerProArgThrGlnAspAla---LeuGlnGlnArgProSerValLeuHis 1879
Db 2353 TCCCT---TCAGTAGAAGTCTCAGGAGCAGCATGTGCAACAGAGACCCAGTGTTCCTCAA 2409
Qy 1880 AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArg 1899
Db 2410 GGAACCAATCGAACCATGTAATCACACTTGGATCCAACTGCTCAGCTACGATCATG 2469
Qy 1900 SerThrSerThrSerSerProVal-----ArgProAlaAlaThrPheProPro 1915
Db 2470 CCAGTGCCTCTGGGGCCCTTCAATAAGCAAGCGCTGCCAGCTCCCGTTACAACACT 2529
Qy 1916 AlaThrHisCysProLeuGlyThrLeuAspGlyValTyrProThrLeuMetGluPro 1935
Db 2530 GCTCGGATGCC---CTGGCTGCTTGTGGATGCTGCAGCTTCTGCCACCCAGATGAT 2586
Qy 1936 ValLeuLeuProLys-----GluAlaProArgValAlaArgProGluArgPro 1951
Db 2587 GTGTCCAAACAAAGAGAGATAGCATGAGCTGCCAGGTTAGAGAAGAAATTTGAGAAGC 2646
Qy 1952 ArgAlaAsp----- 1954
Db 2647 AGTCTACGACAGCTTAGTGAACAGCAGCAGCTAGAGCAGAAAAACCTTGGAGGTGGAGAG 2706
Qy 1955 -----ThrGlyHisAlaPheLeuAlaLysProProAlaArgSer 1967
Db 2707 AGATCTGTTGAGTGTATTACATCTTCTCAGCTTTTCCAAAGTGGCAAGCCC----- 2757
Qy 1968 GlyLeuGluProAlaSerSer-----ProSerLysGlySerGluProArgProLeu 1984
Db 2758 -----CAGCTCATCTTCTCAGTAGTTTATTCTGAGCTGGGAAAGATAAAGGSCCT--- 2808
Qy 1985 ValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn-----Leu 2002
Db 2809 ---CCTCCAAATCCAGATATAGGAAGAGCTAAGGACCCAGAGGGAAGACTTACCTACT 2865
Qy 2003 AlaProHisHisAlaSerProAspProAlaProProAlaSerAlaSerAspProHis 2022
Db 2866 GCAGTAACCTTATAGACGTGATCATCCCGGCAATTCCTCGGACAGGATGCGGAG 2925
Qy 2023 ArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluArgSerLeuGly 2042
Db 2926 GAAGTGGCTCTCAAGTTCAGACTCTTCT-----AGTAGCTTATCT 2967
Qy 2043 TyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerPro 2062

2968 TCTCACAGGTATGAACACACTAGCGATGCTATTAGGTGATTAAGTCTCTGCCAGCTCACT 3027 Db
2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeu 2082 Qy
3028 GCCGCCACCCAGAGAAATCGCAGACTATCAGCAGAGGTGTTAAGCAAAATCAAGCG 3087 Db
2083 GluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHis 2102 Qy
3088 GAAATGATCTCTACCAGACAATATGAAGACCA----- 3120 Db
2103 LeuProHisLeuArgProLeuProGluSerProSerSerProLeuLeuGlnThr 2122 Qy
3121 TTATCATCACTATCGACCA-----CAGCAGAAATCCATCTCTCCCAACAACAGCTG 3171 Db
2123 AlaProGly-----ValLysGlyHisGlnArgValValThr 2134 Qy
3172 CCCCTCTTTCACAGGAGGGAATGGGCAAGTCCAGGACCCATCGGCTGATCACA 3231 Db
2135 LeuAlaGlnHisLeuSerGluValIleThrGlnAspTyrThrArgHisHis----- 2151 Qy
3232 CTTGCTGATCATCTCTCAAAATATCACAAGATTTTCTAGAAATCAAGTTTCTCTCG 3291 Db
2152 -----ProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAla 2169 Qy
3292 CAGACTCCCGCAGCAG-----CCTCTACTTCTACATCCAGAACTCACCTTCTGCT 3342 Db
2170 -----SerCysProValLeuAspLeuArgProProSerAspLeuTyrLeuProPro 2187 Qy
3343 TTGGTATCTACACTGTG-----AGGACTAAACATCAAAACCGTTACAGCCAGAA 3393 Db
2188 ProAsp-----HisGlyAlaPro-----AlaArgGlySerProHis----- 2199 Qy
3394 TCCAGGCTCAGTCTGTCCATCATCAAGACAGGTTCAAGGGTCTCTCCAGAAATCTT 3453 Db
2200 -----SerGluGlyGlyLeuArgSerProGluProAsnLysThrSerValLeuGly 2216 Qy
3454 GTGACAAATTCAGGGGAGTAGGCTCGAAATCCCCAGAGAGGATCAGTCTC----- 3507 Db
2217 GlyGlyGluAspGlyIleGluProValSerProProGluValSerMetThrGluProGlyHis 2236 Qy
3508 ---TCTTCGGAGCCCTACAGCCCATCTCCACCCAG-----GTTCCGGTGTGTCAT 3558 Db
2237 SerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArg 2256 Qy
3559 GAGAAACAGGACGAGTCTGCTCTTGCTCAGAGGGCGCAGAGCTTCAGACAGGAGG 3618 Db
2257 MetGlySerLysSerProGlyAsnThrSerGlnProProAlaPheSerLysLeuThr 2276 Qy
3619 ANTGATGCCGCTCACAGGAGTATAAGTACTTGCCTTCATCTTCAACCAAGCTT--- 3675 Db
2277 GluSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLysLysLeuAsnThr 2296 Qy
3676 GAAATAATACATCCCATGTTAAATCAAGAAGCAGGAGATTTTTCGTAAGTTGAATCC 3735 Db
2297 HisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMet 2316 Qy
3736 TCTGGTGGAGGTGATCTGATATGCGAGTCTGCTCAGCCAGGACCTGATCTTAAATCTG 3795 Db
2317 ProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAla 2336 Qy
3796 CCAGCAGTTACTACGTCAGGCTCAGTTAGCTCTAGAGGCCATCTTTTGTGATCCTGCC 3855 Db
2337 SerThrAsnMetGlyLeuGluAlaIleLeuArgLysAlaLeuMetGlyLysTyrAspGln 2356 Qy
3856 AGT---AATCTTGGGCTGAAGACATTATCAGAAAGGCTCTCATGGAGCTTTGATGAC 3912 Db
2357 TrpGluGlu-----SerProProLeuSerAlaAsnAlaPheAsnPro 2370 Qy
3913 AAGTTGAGGATCATGGAGTTGTGATGTCAGTCCAGGCTATGGGAGTA-----GTGCT 3963 Db
2371 LeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSer 2390 Qy
3964 GGTACTGCCAACACCTCAGTT-----GTGACAGGTGGTGACACAGAGA 4008 Db

2391 AspHisThrLeuThrSerProGlyGlyGlyGly-----LysAlaLysValSerGlyArg 2408 Qy
4009 GAGGAAGGGAGCCCATCACCTCATTAGGAGGAGTTTGCAAAAGCTGATCAGCAAG 4068 Db
2409 ProSerSerArgLysAlaLysSerProAlaPro-----GlyLeuAlaSerGlyAspArg 2426 Qy
4069 TCAACACAGCAGGAAATCTAGTCTCTATACCTGGCAAGGCTACTTTAGGAACGGAACGG 4128 Db
2427 ProProSerValSerSerValHisSerGluClyAspCysAsnArgArgThrProLeuThr 2446 Qy
4129 CCCTCTTCAGTCTCTCTGTATACATTCAAGAGGGATTCACATAGCAGACGCCA----- 4182 Db
2447 AsnArgValTyrGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPro 2466 Qy
4183 GGTGGGGCTGGGAGACAGAGCCCTCTTCAACAGGCTCAACTCAGTTTCTCTATAACCT 4242 Db
2467 LeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAla 2486 Qy
4243 CTGACTATGCGATG-----CTCAGCAGTACTCCACCAACACCCGATTGCGATGCT 4293 Db
2487 GlySerGly---ProLeuAlaGlyProHis-----HisAlaTrpAspGluGluPro 2502 Qy
4294 CCCTCTGGGGTGAACCAAGCAGCTCTCCACACAGACAGATCTGGGACGAGGCT 4353 Db
2503 LysProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517 Qy
4354 GCCCCTGCTCTCAGCACAGTACGAGACCTGTGCGATAGTGAT 4398 Db

RESULT 25
MUSRIP13G LOCUS 4285 bp mRNA linear ROD 10-DEC-2001
DEFINITION Mus musculus RIP-13 (RIP13) mRNA, complete cds.
ACCESSION L78294
VERSION L78294.1 GI:17467267
KEYWORDS RIP13 gene; corepressor; n-cor gene; nuclear hormone receptor.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 4285)
AUTHORS Seol,W., Mahon,M.J., Lee,Y.K. and Moore,D.D.
TITLE Two receptor interacting domains in the nuclear hormone receptor corepressor RIP13/N-Cor
JOURNAL Mol. Endocrinol. 10 (12), 1646-1655 (1996)
MEDLINE 97120602
PUBMED 8961273
COMMENT RIP13 is identical to the N-Cor, except that it has 10 unique amino acids at its N-terminus instead of N-Cor's 1017 amino acids of its N-terminus and contains a 48 amino acids deletion (N-Cor No. 1235-1282). In addition, RIP13 has one more alanine (N-Cor 1589) and a serine deletion (2145) and a amino acid change from A to P. These minor differences may be due to polymorphism. RIP13 sequence has been reconstructed using three clones whose sequences are overlapped partially. There is a deletion form of RIP13 containing a deletion of 119 amino acids (RIP13delta).

FEATURES
Location/Qualifiers
1..4285
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="liver"
1..4285
/gene="RIP13"
1..40
/note="putative"
40..4239
/gene="RIP13"
/note="putative"
/codon_start=1
/product="RIP-13"
/protein_id="AAL40135.1"

Db 1294 CACGAAGCTCCCAAGACGACGAGCTGACCGCGGACTCTATGATGACAGCAGCGCTCGCAGG 1353
Qy 1549 -----ProPheAlaGlyHisLeuProArgGlySerProValThrMetArgLeuPro 1565
Db 1354 ACTCTGTGTGACTACCAAGACCATATCCAGAGGCTCCCGCATGATGAACAGA----- 1407
Qy 1566 ThrProArgLeuGlnGluGlySerLeuSerSerSerLys-----AlaSerGlnAspArgLys 1584
Db 1408 -----ACTCTGTGATGTTCTTCCAGCAAGTCTCCAGTCCCGGAGAA 1452
Qy 1585 LeuThrSerThrProArgGlu-----IleAlaLysSerProHisSerThrVal 1600
Db 1453 TCACCTGTGACCCCAACCCCAAGAGAGATATACCAAGCTCCAGTCCCGGGTG 1512
Qy 1601 ProGluHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGly 1620
Db 1513 GATCCCGCTGTGATCAC-----AGCCCATTTGATCCTCATCACAGGAGTAGCGTGCA 1566
Qy 1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly 1640
Db 1567 GGAGAGGTTTATCGAGGCCACCTACCCACGACTTGGATCCA---GCTATGCCCTTTTAC 1623
Qy 1641 IleProLeuAsp---AlaAlaAlaIleTyrTyrLeuProArgHisLeuAlaProAsnPro 1659
Db 1624 AGGCTTTGGATCTGCGAGCTGCTGCTTACCTGTTACAGAGACAGCTTTTCAACACCCCA 1683
Qy 1660 ThrTyrProHisLeuTyrProTyrTyrLeuIleArgGlyTyrProAspThrAlaAlaLeu 1679
Db 1684 GGATACCCCAAGTACGATCAGCTCAT-----GCAATG 1716
Qy 1680 GluAsn---ArgGlnThrIleLeuAsnAspTyrIleThrSerGlnGlnMetHisAsn 1698
Db 1717 GAGAAATACAGGCAGACAACTCTCAAGATTTACCTTACAGAGACAGCTTTTCAACACCCCA 1776
Qy 1699 ThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSer 1718
Db 1777 CTG-----CGCCTGATGTCACACAGGGAGCTGTCCCCACGAGAGCAG 1818
Qy 1719 SerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValPro 1738
Db 1819 CCACTGGGGCTCCCTTACCCAGCT---ACAAGAGGAATCATTTGACCTGACCAATATGCT 1875
Qy 1739 HisLeuProValLeuValProThrProGlyThrProAlaThrAlaMetAspArgLeu 1758
Db 1876 CCA---ACATCTTAGTGCCTCATCGAGGGGGAACGAGACCCCTCCCTGAGCAGATC 1932
Qy 1759 AlaTyrLeuProThrAlaProGlnProPheSerSerArg---HisSerSerSerProLeu 1777
Db 1933 ACGTATATTCCTGGTACACAGGTTACTTTCCCTCCCGGCCATATACGCTGCTTCTCTG 1992
Qy 1778 SerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGlu 1797
Db 1993 TCTCCAGGACACCCAAACACACTT-----GCAGCAGCTGCAAGTGTGAGAGGGAA 2043
Qy 1798 ArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleLeuThrSer 1817
Db 2044 CGAGAGCGGGAAGAGGAGAGCGGAGAGCTGAGCGCGAGCTGAGCGGTGAACGTGA 2103
Qy 1818 ThrThrThrValGluHisAlaPro-----IleTrpArgProGlyThrGluGlnSer 1834
Db 2104 CGCGAAAGGATCGCTGCTGCTCCCGCTGACCTCTACCTACGACGAGTTTCAAGACAG--- 2160
Qy 1835 SerGlySerSerGlySerSerGlyGlyGlyGlySerSerArgProAlaSerHis 1854
Db 2161 -----CCAGGCGCTCTGGCAGGCCAC 2181
Qy 1855 SerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla---LeuGlnGln 1873
Db 2182 GGATATGTCGCTCCCTTCCCT---TCAGTAAGAACTCAGGAGACCATCTCTGCAACAG 2238
Qy 1874 ArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSer 1893
Db 2239 AGACCCAGTGTTTTCCAGGGCCCAATGGAACCAAGTGTAAATCACACACTTTTGGACCCCACT 2298

Qy 1894 LysProThrValLeuArgSerThrSerThrSerSerProVal-----ArgPro 1909
Db 2299 GCTCAGCTACGATCATGCCACTGCTTCTGGGGCCCTTCCATAGTCAAGGCTCGCCA 2358
Qy 1910 AlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyr 1929
Db 2359 GCCTCCGTTTACACACTGCTCGGATGCC---CTGGGTGCTCTTGTGTGATGCTGCAGCT 2415
Qy 1930 ProThrLeuMetGluProValLeuLeuProLysGlyAlaProArg----- 1944
Db 2416 TCTGCACCCCATGATGGATGTTTCCAAAACAAAGAGAGTAAGCATGAAGTCCAGGTTA 2475
Qy 1945 -----ValAlaAlaArgProGluArgProArgProAlaAlaThrGly 1956
Db 2476 GAAGAAATTTGAGAAGCAGGTCAGCAGCAGTTAGTGAACAGCAGCAGCTAGACAGAA 2535
Qy 1957 HisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSer 1976
Db 2536 AACCTGGAGTGGAGAAGAGATCTGTCAGTGTGTGACACTTCTTTCAGCCCTTCCAAGT 2595
Qy 1977 LysGlySerGluProArgProLeuVal----- 1985
Db 2596 GGCAAGGCCCGCTCATGCTCAGTAGTGTATTCTGAGGCTGGGAAAGATAAAGGCT 2655
Qy 1986 ProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsn-----LeuAla 2003
Db 2656 CTCCAAATCCAGATATGAGGAGAGCTAAGGACCGGAGGAGACTACCATTTACTGCA 2715
Qy 2004 ProHisAlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArg 2023
Db 2716 GCTAACTTCATAGACGTGATCATCCCCGGCAAAATGCTCGACAGATGCGAGGAA 2775
Qy 2024 GluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyr 2043
Db 2776 CGTGGCTTCAAGTTTCAGACTCTTCT-----AGTAGCTTGTCTTCT 2817
Qy 2044 HisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerProSer 2063
Db 2818 CACAGCTATCAACCGCTAGTAGTGCATTGAGTGATAAGTCCCGCAGCTCACCTGCA 2877
Qy 2064 LeuThrHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGlu 2083
Db 2878 CCACCCCGGAAAGACCCACAGGCTATACGCGAGCATGTTAAGCAAAATCAAGCAGAA 2937
Qy 2084 GlyLysLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeu 2103
Db 2938 AATGAGTCCACTCGACAGTATGAAGTCCA-----CTG 2970
Qy 2104 ProHisLeuArg-----ProLeuProGluSerGlnProSerSerSerPro 2118
Db 2971 CATCATTTATCGTCCCGCAGCAGGAATCACCATTCTCCACAGCAACAGCCACCTGCCCA 3030
Qy 2119 LeuLeuGlnThr-----AlaProGlyValLysGlyHisGlnArgValValThrLeuAla 2136
Db 3031 TCTTCCAGTCAAGAGGATGAGCAGGTCGCCAGGACCCATCATCATCATCTGCTGCT 3090
Qy 2137 GlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSer 2156
Db 3091 GACCATCATCTGCAAAATATCACAAAGATTTTGTAGAAAT-----CAAGTTCCTCG 3144
Qy 2157 AlaProLeuProAlaProLeuTyrSerPheProGlyAla-----SerCysProValLeu 2174
Db 3145 CAGCCTTCTACTTCTACATTTCCAAACTTCCACTCTGCTTTGTCATCCACACCTGTA--- 3201
Qy 2175 AspleuArgArgProProSerAspLeuTyrLeuProProAsp----- 2189
Db 3202 -----AGAACTAAACCTCAAGCGCTACAGCCAGATCAACAGTCTCAGACTGCTTG 3255
Qy 2190 HisGlyAlaProAla---ArgGlySerProHisSer-----GluGlyGlyLys 2204
Db 3256 CATCCAGACCAAGTCTCCTAGAGTCTTCTCCAGAAATCTTGTGGATAAATCCCGGGGAAG 3315

CDS

242. >1891
 /gene="NCOR1"
 /codon_start=1
 /product="NCOR1 protein"
 /protein_id="AAH50594.1"
 /db_xref="GI:29792259"
 /db_xref="J0611"
 /db_xref="MIM:600849"
 /translation="MSSSGYPPNNGAFSTQSRYPHPSVQYTFPNTRHQOEAFVDPYR
 SSHLEVSQASGLQQQQOQLRRPSSLSSEFHSQSRPQRRTSYSEFPFPGSPVDHD
 SLESKRLPEQSDSHFORVSAALVPLVPLPEGLRASADAKDPAGFGKHEAPSSPI
 SGQPCGDQNASPSKLSKEELIQSMQDRVDEIAKQEOILKKKKQQQLKEEAAPKE
 PEKPSVPPVQKRSIQVLIYDNRNKAEEAKIPEGLGPKVELPLYNQPSDTKVYH
 ENIKTNQVRKKGLILFFKRNHAKRQEKIQRYDQLEMEWEKKVDRLNNPRKAK
 ESKTREYKEQFEIRKQEQERFORVQGRGAGLSATIAKRSHEHSEIIDLGESEN
 NEKMRQLSVIPMMFIDABQRVRFINNMGLEDPMKVIKDRFMNVWIDHEKFIKFD
 KFIQHPKNGFLIASFLERSKSPDCVLYLYLTKKVNYIKALVRRNYKGRGRNQOQIAR
 PQSEKVEKEEDKAEKTEKKEEKKDBEEKEDSKKKK"
 1553. 1690
 misc_feature
 /gene="NCOR1"
 /note="SANT; Region: SANT SW13, ADA2, N-COR and TFIIB;"
 DNA-binding domains"
 /db_xref="CDD:smart00717"

ORIGIN

Alignment Scores:
 Pred. No.: 9,11e-27 Length: 1891
 Score: 1687.00 Matches: 356
 Percent Similarity: 72.81% Conservative: 59
 Best Local Similarity: 62.46% Mismatches: 103
 Query Match: 12.77% Indels: 52
 DB: 9 Gaps: 12

US-09-522-753-5 (1-2517) x BC050594 (1-1891)

Qy	16	GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaAArgThrHisThr	35
Db	290	CAAGTCGTTATCTCTCTCACTCTGTCCAGTATACATTTCCCAACACCCGCCACACGACG	349
Qy	36	AspValGlyLeuLeuGluTyrGln-----HisHisSerArgAspTyrAlaSerHisLeu	53
Db	350	GAGTTCGCAGTCCTGATTATCTCTCTCTCACTCTTGAAGTCAGTCAGGCATCACGCTT	409
Qy	54	SerProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGln	73
Db	410	TTGCAGCAACAGCAGCAGCAACAGCTTCGAGCGACCTTCCTGCTTTCAGAAATTCAC	469
Qy	74	ProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPro	93
Db	470	CCAGGTTCTCACAGGCCTCAAGAA-----AGGAGAACCTAGTTATGAACCGTTTCATCCA	523
Qy	94	GluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuLeuLeu	113
Db	524	GGCCCATCCCGCAGTGATCATGATTCACCTGGAAATCGAAGCAGCAGCTCGACAGGTT	583
Qy	114	ProAsp-----ProLeuLeuArgProSerPro	122
Db	584	TCGTATTCTCATTTTCAGCGTGTCACTGCTGCGGTTTTCGCTTTAGTCACCCCGCTGCCA	643
Qy	123	LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr	142
Db	644	-----GAAGGCGCTAGGGCT-----TCCTGCAGATGCTAAGAGGATCCAGCATTCGGA	691
Qy	143	GlyLysLeuGlu-----ProValSerProProSerProHisThrAspProGluLeuGlu	161
Db	692	GGCAAAATGAGCTCCATCTCTCCAAATTCGGGGCAACCATGTGTGAGATGATCAAAAT	751
Qy	162	LeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArg	181
Db	752	GCTTCACCTTCAAACTCTCAAGGAGAGATTAATACAGATATGGATCGTGTAGATCGA	811
Qy	182	GluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnLeuGlu	201


```

Qy 561 LysGluLaValAlaSerLysGlyArgLys 570
|||||
Db 1865 AAAGAA-----GACTCCAAAAA 1888

RESULT 27
BC058511 1850 bp mRNA linear PRI 05-NOV-2003
LOCUS Homo sapiens nuclear receptor co-repressor 1, mRNA (cDNA clone
DEFINITION IMAGE:5190073), partial cds.
ACCESSION BC058511
VERSION BC058511.1 GI:35193217
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1850)
Strausberg,R.L., Feingold,B.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Ustin,T.B., Toshituki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.C., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shvechenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriquez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalek,U., Smalios,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1850)
Strausberg,R.
Direct Submission
Submitted (22-SEP-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
REMARK
COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbe-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeeadi, Jacqueline
Schein, Duane Smalios, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 119 Row: d Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 22538460.
Location/Qualifiers
1. .1850

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5190073"
/tissue_type="Colon, Kidney, Stomach, adult, whole pooled"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .>1850
/gene="NCOR1"
/note="synonyms: TRAC1, hn-Cor, N-Cor, hCITS29110, KIAA1047"
/db_xref="LocustID:9611"
/db_xref="MIM:600849"
247. .>1850
/gene="NCOR1"
/codon_start=1
/product="NCOR1 protein"
/protein_id="AAH58511.1"
/db_xref="GI:35193218"
/db_xref="LocustID:9611"
/db_xref="MIM:600849"
/translation="MSSSGYPPNQAFSTQSRYPHSHVQVTPFNTHQHQFAVDDVR
SSHLEVSQASLIQQQQQQLRRRPSLLSEFHPGSDRQERRTSYEFHPGSPVDHD
SLESKRPRLQVSDSHFQVRSAVPLVPLHPEGLRASADAKDPAFGHEAPSPFI
SGQPCGDQNASPSKLSKEBLOSMDRVDIEAKVEQILKLLKKQOOLNPSDPTKVPH
PEKPPPPVPEQKRSIVQIIVYDENRKAKEANKIPEGLGPKVELPLYNQPSDPTKVPH
ENIKTVQMRKLLILFKRNHARKQREQIKQDYDQIMEANEKVKVDRIENNPRKAK
ESKTRYEQVPEIRKQQRQQRQFQGRGAGLSATIAERSEHEISEIIDLSEQEN
NEKMRQLSVIPFMRDAEQRRVKFINMGMEDPMKVYKDRQPMNVTWTHKEIFKD
KFTQHPNKGLIASLYLERKSVDPCLVLYLTCKENYKALVRRNRYKRRGRNQOQIAR
PSOEKVEEKEDEKAEKTEKKKKKK"
1558. .1695
/gene="NCOR1"
/note="SANT Region: SANT SWI3, ADA2, N-Cor and TFIIB"
DNA-binding domains"
/db_xref="CDD:smart00717"

ORIGIN
Alignment Scores:
Pred. No.: 2,67e-26 Length: 1850
Score: 1663.50 Matches: 344
Percent Similarity: 74.77% Conservatives: 56
Best Local Similarity: 64.30% Mismatches: 98
Query Match: 12.59% Indels: 37
DB: 9 Gaps: 9

US-09-522-753-5 (1-2517) x BC058511 (1-1850)
Qy 16 GluProArgTyrProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 295 CAAAGTCGTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 354
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 36 AspValGlyLeuLeuGlyTyrGln-----HisHisSerArgAspTyrAlaSerHisLeu 53
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 355 GAGTTCGAGTCCTCGCTGATTATCGTTCCTCTCTCTCTCTCTCTCTCTCTCTCT 414
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 54 SerProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluPheGln 73
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 TTGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 474
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 74 ProGlyAsnGluArgSerGlnGlnLeuHisLeuArgProGluSerHisSerTyrLeuPro 93
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 475 CCAGGTTCTCAGAGGCTCAGAA-----AGAGAACTAGTATTATGAACCGCTTCATCCA 528
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 94 GluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLeuLeuLeu 113
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 529 GGCCCATCCCGAGTGGATCATGATTCTACTGGAATCGAAGCGACCCACGCTCTGGAACAGGTT 588
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 114 ProAsp-----ProLeuLeuArgProSerPro 122
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 589 TCTGATTCTCATTTTCAGCGGTGTGAGTGTGCGGTGTTTTCCTTTTAGTGCACCCCGCTGCCA 648
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacqueline Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAP Plate: 104 Row: a Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

Location/Qualifiers
1. 3025
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="MGC:56355 IMAGE:5604262"
/tissue_type="Whole body, adult male"
/clone_lib="Sugano SJD adult male"
/lab_host="DH10B"
/note="Vector: pME18S-FL3"
312. 1220
/codon_start=1
/product="Similar to nuclear receptor co-repressor 1"
/protein_id="AAH49302.1"
/db_xref="GI:29179651"
/translation="MSSSGFPPSGSPSSQSRYPHSHVQVTFSTHRPQFVSVDYR
NPLQQRRLSLLSEDFHFKTQVNRKLLIFFRNHRKQKQIKQRYDELMTW
ETHFSRQSDTKVNDHFKTQVNRKLLIFFRNHRKQKQIKQRYDELMTW
EKKVEMNNPRKAKSTREYERQFIRKQEQQRQFQVQQRQGTLSATARS
EHEISEIDGLSEQENNEKQRLSVIPVMYDSQRRVKFINNGLMDDPMKVTYKR
QPMVNVTEHEKEILGRSLCSIPRTSG"

FEATURES
source
Alignment Scores:
Pred. No.: 2,03e-25 Length: 3025
Score: 1628.50 Matches: 404
Percent Similarity: 49.30% Conservative: 127
Best Local Similarity: 37.51% Mismatches: 198
Query Match: 12.32% Indels: 350
DB: 5 Gaps: 28
US-09-522-753-5 (1-2517) x BC049302 (1-3025)
Qy 11 ThrTrpArgAlaThrGluProArgTyrProHisSerLeuSerTyrProValGlnIle 30
Db 345 TCCTTCAGCAGCGAACAGAGCGCTTACCCCCACACTCAGTCAGTACACCTTCAGCTCC 404
Qy 31 AlaArgThrHisThrAspValGlyLeuGluTyrGlnHisSerArgAspTyrAla 50
Db 405 ACACGCCACCCGCGAGGAGTTCTCAGTATCAGCTACCGCAATCCCTCGCAGGAC----- 458
Qy 51 SerHisLeuSerProGlySerIlelleGlnProGlnArgArgProSerLeuLeuSer 70
Db 459 -----CAGCAGCTCGCAGACCTTCACCTACTGTCT 488
Qy 71 GluPheGlnProGlyAsnGluArgSerGlnGluHisLeuArgProGluSerHisSer 90
Db 489 GAATTCACCTCGGACCGAGCGCCAGCGAG-----CCGACAGATGGT 533
Qy 91 TyrLeuProGluLeu-----GlyLysSerGluMetGluPheIleGluSerLys 106
Db 534 TATGAGCAGCAGATTCCACGCCATCTCAGCTCAACAGGAGCAGAGGCTCTCGAGAGCAAG 593
Qy 107 ArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThr 126
Db 594 CGTCCCGCATAGAG----- 608
Qy 127 GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGlu 146
Db 608 ----- 608

Qy 462 ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLeuAsnGluAsnTyrLysSer 481
Db 1593 AGTGTCTCGATGCTGTTTGTATTAATTAACCAAGAAATGAGAATATAAGCC 1652

Qy 482 LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGln 500
Db 1653 CTCCTCAGAAGGAATTATGGAACACGAGGAGGAGAAACAGCAA----- 1697

Qy 501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1698 -----ATTGCTGAGCCCTCGCAAGAGAGAAA 1724

Qy 521 AspGluLysGluLys 525
Db 1725 GTAAATAAATAAATAA 1739

RESULT 29
BC049302
LOCUS
DEFINITION
BC049302 3025 bp mRNA linear VRT 07-OCT-2003
clone MGC:56355 IMAGE:5604262, complete cds.
ACCESSION
BC049302.1 GI:29179650
VERSION
MGC.
KEYWORDS
MGC.
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

REFERENCE
AUTHORS
1. (bases 1 to 3025)
Klausner, R.D., Collins, F.S., Wagner, L., Shemmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Narusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 3025)
Strausberg, R.

Direct Submission
Submitted (24-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Dr. Sumio Sugano
CDNA Library Preparation: Dr. Sumio Sugano
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
CDNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Varon Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo

```
Qy 147 ProValSerProProSerProProHisThrAspProGluLeuLeuValProProArg 166
Db 608 -----
Qy 167 LeuSerLysGluGluLeuLeuGlnAsnMetAspArgValAspArgGluLeuThrMetVal 186
Db 609 -----
Qy 187 GluGlnGlnIleSerLysLeuLysLysGlnGlnLeuGluGluGluAlaAlaLys 206
Db 618 GAAACACACTTTTCAAGA-----
Qy 207 ProProGluProGluLysProValSerProProLeuSerLysHisArgSerLeu 226
Db 635 -----
Qy 227 ValGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGlu 246
Db 635 -----
Qy 247 GlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyr 266
Db 636 -----
Qy 267 HisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArg 286
Db 663 CATGACAAACATTAAAGCAATCAGGTGATGAGGAAGAAGCTGATTCTGTTTTTCAAGAGG 722
Qy 287 ArgAsnHisAlaArgLysGlnTyrLysGlnLysPheCysGlnArgTyrAspGlnLeuMet 306
Db 723 AGAAATCATGCACCAACAGAGGAGGAGCAGAAATTTGTGACGGGTATGATGAACACTATG 782
Qy 307 GluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArgArgAlaLysGlu 326
Db 783 ACAGAGTGGGAGAAAAGTTGACGTATGGAAAACACCCACGCCGTAAAGCCAAAGAG 842
Qy 327 SerLysValArgGluTyrTyrGluLysGlnPheProGluLeuArgLysGlnArgGluLeu 346
Db 843 AGCGGCACACGAGAAATACTACGAGAGGCGAGTTTCTGAAATCCGCAAAACAGCGCGCA 902
Qy 347 GlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAla 366
Db 903 CAGAGCGTTTCCAG---AGAGTTGGCCACAGGGGGACAGCCCTGTCTGCCACCATCGCA 959
Qy 367 ArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGlu 386
Db 960 AGAAGTGAAATGATGATTTCTGAGATCATTTGACGGCTCTCGAGCGAGGAGAACATGAG 1019
Qy 387 LysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnGlnArg 406
Db 1020 AAGCAGATGAGCGAGTTGTCCGTCAATTCCTCCCGTGATGATGATTCAGAGCAGCGGAGA 1079
Qy 407 IleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArg 426
Db 1080 GTAAGGTTTATATATGAAACGGTCTGATGATGACCCCATGAAAGGTCTCAAGTCCCGC 1139
Qy 427 GlnValMetAsnMetTyrSerGluGlnGluLysGluThrPheArgGluLysPheMetGln 446
Db 1140 CAGTTCTGATGATGCTCGACCGAAACACAGAGAGAGAT-CTTAGGGAGAAATTTGTGCAG 1198
Qy 447 HisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCys 466
Db 1199 CATCCCAAGAACTTCGGGCTTAATCGCTCTTTCTGAGAGGAGAGTGTGTCTGATTGT 1258
Qy 467 ValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSer 486
Db 1259 GTGCTGTACTACTATTGTGACTAAGAGAGTCAAACTATATAAACGCTTGTGACAGCAAT 1318
Qy 487 TyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506
Db 1319 TTTGGAAACCGAGGCCGCGAATCAGCAGATACACAGCCCTTCACAGGAGAGAGACACA 1378
Qy 507 GlnGlnGlnGlnProMetProArgSerSerGlnGluLysAspGluLysGluLysGlu 526
```

```
Db 1379 GAAGACAAGAACCAAGAGGACAAATCCGAGAAAACCGAGAAAGAGGACGAGGAGAAA 1438
Qy 527 LysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLeuLeu 546
Db 1439 AAGGAGAGAGAGAGAGAAAGATGAAAA-----GAGGACTCTAAGGACATCATG 1486
Qy 547 LysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSer 566
Db 1487 AAAGATAAA---GAGAAAGTGTGAGGAGAGAGAGAGAGGTTAAGGAGCAGACACCA 1543
Qy 567 LysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIle---ThrArgSer 585
Db 1544 CGTGCGCGAAAGACCGCAACAGCAGCGCGCGCGCAAGGCGCATCACCATCGCTCC 1603
Qy 586 MetAlaAsn----- 588
Db 1604 ATGCCCAATTAATGAGGCTGCATCTGTGTCTGTAAGACCCACCTCCACACCACTGAA 1663
Qy 589 -----GluAlaAsnSerGluAlaIle 596
Db 1664 CCAGTTTTTGCAGAGCCAGCACCCCTTCCAAACCTTGAGCGCAGACACAGAAACCACTCGA 1723
Qy 597 ThrProGlnGlnSerAlaGluLeuAlaSer-----MetGluLeuAsn 610
Db 1724 GAGCCCGCAAAACCTCTATTTCAGCCAGCGCAGACACAGAGCGCGGTTGAAGCTGGA 1783
Qy 611 GluSerSerArgTyrThrGluGluMetGluThrAlaLysLysGlyLeuGluHis 630
Db 1784 GAGACGTCTCGTGGAGGAGGAGATGAGATCGCCAAAAGGTTCTGGTTGAACAT 1843
Qy 631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
Db 1844 GSCAGGAACCTGGTCAGCCATTGCTAAGATGGTAGGAGTAAGAGTGAAGCTCAATGTAAG 1903
Qy 651 AsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnHisLys 670
Db 1904 AACTTTTACTTCAACAAAGAGCGCATACAACTAGACAGTCTTACTACGCAACACAAG 1963
Qy 671 LeuLysMetGluLysGluArgAsnAlaArgLysLysLysAlaProAlaAlaAla 690
Db 1964 -----GGTTCCCGCGGCTTCTGTGAAATCGAGACCCCTCGAG 2002
Qy 691 SerGluGluAlaAlaPheProProValValGluAspGluGluMetGluAlaSerGlyVal 710
Db 2003 AGTGAT-----CCAGTGCCTGATGACCGATGAG----- 2029
Qy 711 SerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGlu 730
Db 2029 ----- 2029
Qy 731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSer 750
Db 2030 -----GAGAGTCGCGTCCA-----GATAACAGTTCCAGACACAGAGAGT 2068
Qy 751 IleProSerPro-----HisThrGlu 757
Db 2069 CGGCTTCTCCAACCTCAAACTGATCCCTCCAAACCTCAGAACTCACTAATCATACCAGT 2128
Qy 758 AlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGly 777
Db 2129 TCAGCAGTGTGATGCGGATCAGAC-----CAGGACGCTCACTGCGACCAAGGC 2176
Qy 778 Pro-----ProProGlyProProThrProThrProArgArgThrSerArgAlaProIle 794
Db 2177 CCGTGTCAGATGCCAAAGCCCGAGGATCTCCATACCGTGAGCTG---AGGTTAAATTT 2233
Qy 795 GluProThrProAlaSerGluAlaThrGlyAlaProThrProProAlaProProSer 814
Db 2234 GAGAAGAGTCCAGAGGAGGAGCC----- 2257
Qy 815 ProSerAlaProProProValValProLysGluGluLysGluGluThrAlaAlaAla 834
```



```
* 24356 24455: gap of 100 bp
* 24456 26757: contig of 2302 bp in length
* 26758 26857: gap of 100 bp
* 26858 28317: contig of 1460 bp in length
* 28318 28417: gap of 100 bp
* 28418 30781: contig of 2364 bp in length
* 30782 30881: gap of 100 bp
* 30882 32789: contig of 1908 bp in length
* 32790 32889: gap of 100 bp
* 32890 35214: contig of 2325 bp in length
* 35215 35314: gap of 100 bp
* 35315 37181: contig of 1867 bp in length
* 37182 37281: gap of 100 bp
* 37282 39642: contig of 2361 bp in length
* 39643 39742: gap of 100 bp
* 39743 41978: contig of 2236 bp in length
* 41979 42078: gap of 100 bp
* 42079 44963: contig of 2885 bp in length
* 44964 45063: gap of 100 bp
* 45064 47265: contig of 2202 bp in length
* 47266 47365: gap of 100 bp
* 47366 49847: contig of 2482 bp in length
* 49848 49947: gap of 100 bp
* 49948 52474: contig of 2527 bp in length
* 52475 52574: gap of 100 bp
* 52575 54767: contig of 2193 bp in length
* 54768 54867: gap of 100 bp
* 54868 57938: contig of 3071 bp in length
* 57939 58038: gap of 100 bp
* 58039 60538: contig of 2500 bp in length
* 60539 60638: gap of 100 bp
* 60639 63422: contig of 2784 bp in length
* 63423 63522: gap of 100 bp
* 63523 66038: contig of 2516 bp in length
* 66039 66138: gap of 100 bp
* 66139 69189: contig of 3047 bp in length
* 69189 69288: gap of 100 bp
* 69289 72321: contig of 3036 bp in length
* 72322 72421: gap of 100 bp
* 72422 75240: contig of 2819 bp in length
* 75241 75340: gap of 100 bp
* 75341 77598: contig of 2258 bp in length
* 77599 77698: gap of 100 bp
* 77699 82485: contig of 4787 bp in length
* 82486 82585: gap of 100 bp
* 82586 86177: contig of 3592 bp in length
* 86178 86277: gap of 100 bp
* 86278 91129: contig of 4852 bp in length
* 91130 91229: gap of 100 bp
* 91230 94475: contig of 3246 bp in length
* 94476 94575: gap of 100 bp
* 94576 98382: contig of 3807 bp in length
* 98383 98482: gap of 100 bp
* 98483 103891: contig of 5409 bp in length
* 103892 103991: gap of 100 bp
* 103992 109749: contig of 5758 bp in length
* 109750 109849: gap of 100 bp
* 109850 116075: contig of 6226 bp in length
* 116076 116175: gap of 100 bp
* 116176 122437: contig of 6262 bp in length
* 122438 122537: gap of 100 bp
* 122538 132392: contig of 9855 bp in length
* 132393 132492: gap of 100 bp
* 132493 141881: contig of 9389 bp in length
* 141882 141981: gap of 100 bp
* 141982 161970: contig of 19989 bp in length.
```

FEATURES
source

```
1. .161970
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12"
```

```
/clone="RP11-665C13"
/clone_lib="RPC1-11 Human Male BAC"
1. .1052
/note="assembly_fragment"
1153. .2378
/note="assembly_fragment"
2479. .3812
/note="assembly_fragment"
3913. .5368
/note="assembly_fragment"
5469. .6519
/note="assembly_fragment"
6620. .7852
/note="assembly_fragment"
7953. .8464
/note="assembly_fragment"
clone_end:SP6
vector_side:left
8565. .10347
/note="assembly_fragment"
10448. .12145
/note="assembly_fragment"
12246. .14006
/note="assembly_fragment"
14107. .15358
/note="assembly_fragment"

Alignment Scores:
Pred. No.: 2.52e-19 Length: 161970
Score: 1396.00 Matches: 529
Percent Similarity: 32.58% Conservative: 104
Best Local Similarity: 27.23% Mismatches: 442
Query Match: 10.56% Indels: 872
DB: 2 Gaps: 52

US-09-522-753-5 (1-2517) x AC027706 (1-161970)
Qy 706 GluAlaSerGlyValSerGly-----AsnGluGluGluMetValGluGluAlaGluAla 723
Db 60122 GAGCCTCGGGGCTGTTTGGCATCAGCGCGACAGAGCGGCTCATCCCGAGGTCAGGGCC 60181
Qy 724 LeuHis-AlaSerGlyAsnGluValProArgGlyGluCysSerGlyPro----- 739
Db 60182 ATGCATCTCTTCTGGAGGC-----CCCGCAGACCCCTGTCTCATC-CCCTCTTCTCTC 60234
Qy 740 -----AlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThr 756
Db 60235 GCCTTCGACAGCCACTGTCAACACAGCTCAGACCCAGAGCATCCCTCTCTCTCACAC 60294
Qy 756 rGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAs 776
Db 60295 TGAGGCGCCACAGACACAGCGGCGAGAATGGGCCCAAGCCCCAGCCACCTGGGCGCGGA 60354
Qy 776 pGlyProProGlyProProProProProPro-ArgArgThrSerArgAlaProIleGluP 796
Db 60355 CGGGCCACCCCGAGGCGCCACCCACCCAGAGGAGATCCCGGCC-CCCACTGAGC 60413
Qy 796 roThrProAlaSerGluAlaThrGlyAlaProThrProProProProAlaProSerProS 816
Db 60414 CCACCCCGGCTCTGAAGCCACCGAGCCCTCAGCCCCCAGAGAGAGAGAGAGAGAGAG 60473
Qy 816 erAlaProProValProLysGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 836
Db 60474 CTGCACCTCTCTGTGTGTCCTCCCAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 60533
Qy 836 roValGluGluGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 855
Db 60534 CATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60593
Qy 856 -----G 856
Db 60594 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60653
Qy 856 lyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGluGluGluGlu 875
```


Qy	1411	sGluGlyLeuValAlaThrValIysGluAlaGlyArgSerIleHisGluIleProArgG1	1431	Qy	1682	ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThr	1701
Db	62660	TGTCAGGTGTTCCACACTCATGTGGAGGATGTTAGGTGTTTCACACTCATGGCAAGAGTA	62719	Db	63602	CGCGAGACCATCATCAATGACTACATCATCTCGCAGCAGATGCACCAACGCGCCAC	63661
Qy	1431	uGluLeuArg-----HisThrProGluLeuProLeuAlaProAr	1444	Qy	1702	AlaMetAlaGlnAtqAlaAspMetLeuArgGlyLeuSerProAtqGluSerLeuAla	1721
Db	62720	TCACATATACACAGGGTGAGGGTGTGGGTATTCACACCCATGTGTGAGGTATTCACACT	62779	Db	63662	GCATGGCCCGAGCGAGCTGATATGCTGAGGGCCCTCTCGCCCGCGAGTCTCTCGCTGGCA	63721
Qy	1444	gProLeuIleGluSerIleThrGlnGlyThrProLeuIleTyrAspThrGlyAlaSe	1464	Qy	1722	LeuAsnTyrAlaAlaGlyProArg-----	1729
Db	62780	CGTCAGGATGTTAGTGTTCACACTCATGTGCACAGAGGTGCACATATACACAGGGTGAGG	62839	Db	63722	CTCAACTAGCTCGCGGTCCCGAGGTGAGTGGGTGGCGAGCACACCTCCGCTGGGTTTG	63781
Qy	1464	r-----	1464	Qy	1729	-----	1729
Db	62840	TGTTGGGTATTAAACACCCATGTGTGAGGTGTTTCACACTCATGTGTGAGGATGTTAGGTGT	62899	Db	63782	GCCTTTATTTCCAAAGACATGGCGTGCCTCTGTGGCCCTCGCGGAGGAGCTAGACCTGG	63841
Qy	1465	-----ThrThrGlySerIleHisAspValArgSerLeuIleGly-----	1478	Qy	1729	-----	1729
Db	62900	CACACTCATGGCAAGAGTATCACATATACACAGGGTGAGG---GTGTGGGTATTCACAC	62956	Db	63842	TCACCTTGTGGGTCACTTGTGTGTAACCGAACCTGAGTGGGTGGCCTGGGGTTGTGCGTGC	63901
Qy	1479	-----SerProGlyArgThrPheProProVa	1487	Qy	1729	-----	1729
Db	62957	CCATAGTGAAGTGTCTAGTATTACACACCCGTGGAGTCCA-----GAATGGCCATCATC	63010	Db	63902	TGTGGGTGCTGCTTGGCATCTGTTAGTAGTGAGTGACACAGCGTGTGGCTCTGCTGCATCC	63961
Qy	1487	lHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu-----Ar	1501	Qy	1729	-----	1729
Db	63011	TCCTAAATTGATATTCTAATGTATCTCCCTGCCCATAGAGATCATGCCGGGATGTCT	63070	Db	63962	TCAGTGGGTGTGCGTGTCATCTGTATATATCTTCTAGGATACAGGGGCGCTCAGGAGTTTAA	64021
Qy	1501	gAlaCysTyrGluLeuSerLeuIysSerArgProGlyThrAlaSerSerSerGlyIleSe	1521	Qy	1729	-----	1729
Db	63071	GGTGTGTTTGAATACAGGAGGAGAGAGAGAACTCTCCAGAGCTGTTCAGGTGCCTC	63130	Db	64022	GATCAAAATGTGGCCGGGCAACAGTGGTCTCATGCTGTATATCCACGACTTGGGGAGGGCG	64081
Qy	1521	rIleAlaArg-----GlyAlaProValIleValProGluLeuGlyLysPr	1536	Qy	1729	-----	1729
Db	63131	TTGGTTCAGTCTACACACTCTTGGGTGCGCTGTCTGTGTCCGAGCATCCCTTCT	63190	Db	64082	AGGCAGGTGGATACAAAGGTCAAGATTTCGAGACCAGTCTGACCAACATGTTGTAACCCG	64141
Qy	1536	oArgGlnSerProLeuThrTyrGlu-AspHisGlyAlaProPheAlaGlyHisLeu-Pro	1555	Qy	1729	-----	1729
Db	63191	GGAGCAGAGTGA---ACCTGGGAATGGCCACAGTCTTCTCACTATTGACCCAGTCACCT	63247	Db	64142	TCTCTCTAAAAATACAAAAAATTAGCCAGGCATGGTGTATGCGCACCTGTGGTCCACGCTA	64201
Qy	1556	ArgGlySerPro---ValThrMetArgGluProThrProArgLeuGlnGlySerLeu	1574	Qy	1729	-----	1729
Db	63248	CGGCTTGTCCCTTGGTGCCAGTGGCTGGCCCTCCCTCCCTCATTT-----	63292	Db	64202	CTCACTAGGCTGAGGCAAGAGAAATCATCTTGAAACCCAGGAGTGGAGTTACAGTGAGTGG	64261
Qy	1575	SerSerSerIysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLys	1594	Qy	1729	-----	1729
Db	63293	-----ACC	63295	Db	64262	AGATTTTACCATTGCACCTCCAGCTGGGCAACAGGGCAAGACTGTGTCTCGAAAAAANA	64321
Qy	1595	SerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeu	1614	Qy	1729	-----	1729
Db	63296	TGCCCTTTCGTATGCTGCCACAGCAGCAGCGGTGCCCC-----	63334	Db	64322	AAAGATGAAAAATGTAGGCTGTTTGGAGTTTGTTCCTTTGTTTAAACAGGCCACAG	64381
Qy	1615	LeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPro	1634	Qy	1729	-----	1729
Db	63335	-----TCCACACTCCCCCAAGCCTGCACACC	63361	Db	64382	CTGCTTTCGTCACACGTTCCAGGGCCATCCTCAGAAATGCTTCTGGAATAACCAAGTT	64441
Qy	1635	ThrSerIleProArgGlyIle-----	1641	Qy	1729	-----	1729
Db	63362	CACCTGGTGCCGTCCTCGGTGCAGTGCAGTCCCGCAGGACGAGATTGGGTGGCTGAGC	63421	Db	64442	CTAGCTGGGGCTCAGCTGGAAAAAGCTGNAATGCACACTTAAGTATTTTGAACAGTGAGGAT	64501
Qy	1641	-----	1641	Qy	1729	-----	1729
Db	63422	TNN	63481	Db	64502	TGAATACAGGGAATGGATTGTATAGCGCTCAGAGGCTGAAGGGGCACAGAGGGCTTGAGA	64561
Qy	1642	ProLeuAspAlaAlaAlaTyrTyrLeuProArgHisIleuAlaProAsnProThrTyr	1661	Qy	1729	-----	1729
Db	63482	NN	63541	Db	64562	TGGGAACAGTGAGGGCAGCTGCAGGAGATGCCCGGCTCGGGCTTGGGAGCAGAAGAGG	64621
Qy	1662	ProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn	1681	Qy	1729	-----	1729
Db	63542	CCGACCTGTACCACCTACCTACCTCCGGGTACCCCGACACAGCGGCGCTGGAGAAC	63601	Db	64622	AGGTGGTACCAGAGAACCTGAGCATTCAGAAAAAGGGTTCCATGGCGGCTCTCGGAGCC	64681
Qy	1729	-----	1729	Qy	1729	-----	1729

```
Db 64682 GAGGAGGAGTGCCTGCCACCACTCTGCTGGCTCCAGGAGTGTGTCGCTCTCCAG 64741
Qy 1729 ----- 1729
Db 64742 GAGGGTATCTGGCCGCTGGGAGCCTGGCCCTCTCTCCCTGTGGCTACAGCCCTGGC 64801
Qy 1730 -----GlytIleAspLeuSerGlnValProHisLeuProValLeuVal 1744
Db 64802 CCAACACCTCCGAGGCAATCATGACCTCTCCAAAGTGCACACCTGCTGTGCTCGTG 64861
Qy 1745 ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaValLeuProThrAla 1764
Db 64862 CCCCCGACACAGCACCCCGACCGCCATGAGCGCGCTTGCCTACCTCCCAACCGCG 64921
Qy 1765 ProGlnProPheSerSerArgHisSerSerProLeuSer----- 1778
Db 64922 CCCCAGCCCTTCAGACGCCGCCACAGCAGCTCCGCCACTCTCCCAAGTAGCGCACTGCC 64981
Qy 1778 ----- 1778
Db 64982 CAGTCTGGGGGTGGGACCCCGGCATCCATGGAGCGCGCTGGGGATGGGGCGGCGAGA 65041
Qy 1778 ----- 1778
Db 65042 AGCCTGCTCTCTTTCCACCCCGAAGACAAAGCCAGGCTCTTCTCGGCCCTGGGGCT 65101
Qy 1779 -----Pro-GlyGlyProThrHisLeuThrLysProThrThr 1791
Db 65102 GAGTCTCTGGCCTTTGGGTTTCTTAGGAGGTCCAAACACACTTGACAAAACCAACCCAC 65161
Qy 1791 rSerSerSerGluArgGluArgAspArgGluArgGluArgAspArgGluArgGluArgG 1811
Db 65162 GTCTCTGTCGAGCGGAGGAGACCGGGATCGAGACGGGACCGGATCGGGAGCGGA 65221
Qy 1811 uLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleThrParg----- 1828
Db 65222 AAGTCCATCTCTACGTCCTCCACGACGCGGTGGAGCAGCCACCATCTGGAG-ACCTGGTA 65280
Qy 1828 ----- 1828
Db 65281 GGGCATCAGACCCCCACCCCGCTCCGGGACTCTTGTGGCGCGCAAGAGGCTCCCC 65340
Qy 1829 -----ProGlyThrGluGlnSerSerGlySerSerGlySerSe 1841
Db 65341 CTGCTGATGCACATGACTGTACACAGGTACAGACGAGCAGCGGCGAGCGGCGAGCAG 65400
Qy 1841 rGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSe 1861
Db 65401 CGCGGGGGTGGGGCAGCAGCAGCGCGCCGCTCCCACTCCCATGCCACCAAGCACTC 65460
Qy 1861 rProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnth 1881
Db 65461 GCCCATCTCCCTCGGACCCAGAGTGCCTCCAGCAGAGACCCAGTGTGCTTCAACAAC 65520
Qy 1881 rGlyMetLysGlyIleThrAlaValGluProSerLysProThrValLeuArg 1899
Db 65521 AGGCATGAGGGTATCATCACCGCTGTGGAGCCCGACGACCGCCACGCTCTCTGAGG 65575

RESULT 31
AK127788
LOCUS AK127788
DEFINITION Homo sapiens cDNA FLJ45889 fig, clone OCBF3022166, moderately
similar to Nuclear receptor co-repressor 2.
ACCESSION AK127788
VERSION AK127788.1 GI:34534855
KEYWORDS oligo capping; fig (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
```

```
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,Y., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yanashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahara,K., Masuo,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2914)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1..2914
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="OCBF3022166"
/tissue_type="brain"
/clone_lib="OCBF3"
/dev_stage="fetal"
/note="cloning vector: pME18SFL3"
1848..2273
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:34534856"
/translation="MTEPGHSASVYPLLRYDGEOTEPGRMGSKSGNTSQPPAFPSK
LTESAMVKSKEINKLNTNRNEPEYKGPAGSGMGSPQVGWPAWAAGAAQW
QKPTVHLKLSGPADHLPPGLPHIWAGPQGLDLLGFP"

CDS
1848..2273
/note="unnamed protein product"

ORIGIN
Alignment Scores:
Pred. No.: 1,25e-20 Length: 2914
Score: 1391.00 Matches: 334
Percent Similarity: 49.65% Conservative: 24
Best Local Similarity: 46.32% Mismatches: 94
Query Match: 10.53% Indels: 269
DB: 9 Gaps: 13
US-09-522-753-5 (1-2517) x AK127788 (1-2914)
Qy 1823 HisAlaProIleThrParg---ProGlyThrGluGlnSerSerGlySerSergly----- 1839
Db 109 TAGCCCCCACCCTGCCCGCTGCTGTGCGACTGGGGAGTGGAGGACGCGG 168
Qy 1840 -----SerSerGlyGlyGlyGlyGly 1846
Db 169 TGAAGAGAGAGTGTAGCGGGGCGAGGACCTCGCTGTGCGAGGAGGAGTGGGTCTGGC 228
Qy 1847 SerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg 1866
Db 229 TCTTGCATCTTGCCTGTCTCC-----CAGCCCTGTATCCCCAGA 267
Qy 1867 ThrGlnAspAlaLeu-GlnGlnArgPro-----SerValLeuHi 1879
Db 268 AAAAGGGGAGCCCTCTGCTCTGAGCCCTCGCTTGCCTAGTTCATGCTCTCTCTG 327
Qy 1879 sAsnThrGlyMetLysGlyIleThrAlaValGluProSerLysProThrVal-LeuA 1899
Db 328 TTGACTTGGATGGCGAGCTATAGCCAGGCGGGCCCGGGGACCCAGGTCATC- 386
Qy 1899 rgSerThrSerThrSerSerProValArgProAlaThrPheProAlaThrHisC 1919
```


Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES		Location/Qualifiers	
source	1..3575	/organism="Homo sapiens"	
	1..3575	/mol_type="mRNA"	
gene	1..3575	/db_xref="taxon:9606"	
	1..3575	/clone="DKFPz434M075"	
CDS	1..3575	/tissue_type="testis"	
	1..3575	/clone_lib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI"	
CDS	1..3575	/dev_stage="adult"	
	1..3575	/gene="DKFPz434M075"	
CDS	1..3575	/note="nuclear receptor co-repressor, partly"	
	1..3575	/codon_start=2	
CDS	1..3575	/product="hypothetical protein"	
	1..3575	/protein_id="CAB70854.1"	
CDS	1..3575	/db_xref="GI:680844"	
	1..3575	/db_xref="SPTREMBL:O9NSZ0"	
CDS	1..3575	/translation="SQSAIKHNKVKSLITGPKSLSRGMPLEIVPENIKVGRGKYED VKAGETVRSRHSIVSVSGPVSRLASTLHEAPKAQLSPGIYDDTSARETPVSYONTMSRG SPMMNRSDVTISNSKTNHERKSTLPTQRESIPAKSPVPGVDVPSVSPDPRHRQ STAGEVRSRLHPLDPAFPFHALDPAAYALFQRLSPTPGPGPSQYQIYAWENTRQ TILNDYITSOQMVPNDPVARGLSPREQPLGLFPATRGIIIDLTNMPPTILVPHPGG TSPPMDRIYIPCTOITPFPVNSASMPGPHTHLAAAASAEEREREREKERERE RIAAASDLVLRGSEOPGRGSHGVRSVPSVTOETMLQORPSVFOGTNGTSVIT PLDTAQLRIMPLPAGPSPISQGLPASRYNTADALAALVDAASAPQMDVSKTKESK HEARLEENRSRAVSEQQLEKLEVEKSVQCLYTSSTSPFSGKQPHSVVYS EAGKDGPPKPSYEELRTGKTIITAFNIDIVIITRQIASDKDAREGQSDDSS SLGSHRYETSDAIEVSPASSPAPPOEKLTQYQVEVVASQNAENDPTREYGLPHY RPOCESPOQLPPSQAGMGQVPRTHLITLADHCQIITQDFARNVSSQTPOQ PPTSTFONSALVSTVPTKTSNRYSPESQAOVHORGPSRVSPENLVKSGRSP GKSPERHVSSEPEPLSPPOVPVHESKODLLLSORGAEPAEONDRSPGSIYLV PSFTKLENTSPWKSXKQIFKRLNSGGSDSDMAAOPGTEIFNLPAVTTSGSVS RGHSPADPAGNLGLIEDIIRKALMGSPDKVEDHGVNMSQPMGVFGTANTSVTSGE T RREEDPSGSHGVCKPKLISKNSRKSQPIPGQYGLTGERPSSVSSVSHSGDYHRQ TPGWAWEDRPSSTGSGTGPYFNPLTMRSLSTPTPIACAPSANQNAAPHQONRIWERE PAPLLSAQYETLSDD"	
polyA_site	3511	/gene="DKFPz434M075"	
	3511	/gene="DKFPz434M075"	
ORIGIN	3.77e-18	Length: 3575	
	1272.00	Matches: 401	
Alignment Scores:	48.03%	Conservative: 149	
	35.02%	Mismatch: 202	
Best Local Similarity:	9.63%	Indels: 54	
	9	Gaps: 54	
Query Match:	9	Gaps: 54	
	9	Gaps: 54	
DB:	9	Gaps: 54	
	9	Gaps: 54	
US-09-522-753-5 (1-2517) x HSM802410 (1-3575)	1470	LysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisPro 1489	
	20	AAACAACAATGTCATACTTAAATACGGGGGCTAGCAAACTATCCCTGGGAATGCTCCG 79	
QY	1490	LeuAspValMetAlaSer---AlaArgAlaLeuGluArgAlaCysTyrGlu----- 1505	
	80	CTGGAAATGTCAGAGACATATAAGTGTAGAACGGGGAATAATATGAGATGTGAAA 139	
DB	1506	-----GluSerLeuIleSerArgProGlyThrAlaSerSerSerGlyCysSerIleAla 1523	
	140	GCAGGCGAGACCGTGGTCCCGGACACGTCAGTGGTGAAGCTCTGGCCCTCGCTTT 199	
QY	1524	ArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyr 1543	
	200	AGTCCACA-----CTGATGAAGTCCCAAGACACACTGAGCCCTGGATTAT 250	
DB	1544	GluAspHisGlyAla-----ProPheAlaGlyHisLeuProArgGlySer 1558	
	251	GATGACACCATGTCAGGAGGAGCCCTGTGTGATTATCAAAACACCATGTCAGAGGTCA 310	
QY	1559	ProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerIle 1578	
	311	CCCATG-----ATGAACAGAACTTCTGATGTTTCAATTTCTTCAACAAG 355	
DB	1579	Ala---SerGlnAspArgLysLeuThrSerProArgGlu-----IleAla 1593	
	356	TCTACCAATCATGAAGGAATTCGACACTGACCCCTACCCAGAGGAAAGTATCCACGG 415	
QY	1594	LysSerProHisSerThrValProGluHisProHisProHisProHisProHisProHis 1613	
	416	RAGTCTCCAGTGCCTGGGTGGACCTCGTGAGCCAC-----AGTCCGTTTGATCC 469	
DB	1614	LeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAsp 1633	
	470	CATCACAGAGGAGCAGCTGCGAGGCGAGGTTCATCGAGCCACCTCCACCGACTTGAT 529	
QY	1634	ProThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuPro 1652	
	530	CCA---GCCATGCCCTTTTACAGGCTTTGGATCCTGAGGGCTGCTTACCTGTTTTCAG 586	
DB	1653	ArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGly 1672	
	587	AGACAGCTTTCACCAACTCCAGTTTACCACTAGTATCAGCTTTTAC----- 634	
QY	1673	TyrProAspThrAlaAlaLeuGluAsn---ArgGlnThrIleIleAsnAspTyrIleThr 1691	
	635	-----GCAATGGAGAACACACACAGACACATCTTAAATGATTACATTACC 679	
DB	1692	SerGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg 1711	
	680	TCACAACATGCAAGTGAAGTTG-----CGTCCAGATGTGCGCCAG 721	
QY	1712	GlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIle 1731	
	722	GGAATCTCCCAAGAGAGAGCAGCAGCTGGGTCTCCCATCCCAAGCA---ACGAGAGGAAATC 778	
DB	1732	IleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrPro 1751	
	779	ATTGACCTGACCAATATGCTCCA---ACAATTTAGTGCCTCATCCAGGGGGAACAAGC 835	
QY	1752	AlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArg 1771	
	836	ACTCTCTCCCATGACAGAACTCACTTATTTCTGTTACACAGATTACTTCTCCCTCCAGG 895	
DB	1772	HisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThr 1790	
	896	CCGTACAACCTCTCTTCCATGTCTCCAGGACACCCCAACACACTT-----GCAGCT 946	
QY	1791	ThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArg 1810	
	947	CTGCAAGTCTGAGAGGGAACGGGAACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1006	
DB	1811	GluLysSerIleLeuThrSerThrThrValGluHisAlaProIleThrArgProGly 1830	
	1007	-----ATTGTCGAGCTTCTCCGACCTC-----TACCTGCGCGCCAGGC 1045	
QY	1831	ThrGlnSerSerGlySerSerGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1850	
	1046	TCAGAACAG-----CCTGGCCGA 1063	
DB	1851	ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAla 1870	
	1064	CCTGGCAGTCATGATATGTTGCTCCCTTCCCT---TCAGTAGAAGTCTAGGAGACC 1120	
QY	1871	LeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAla 1889	
	1121	ATGTTGCAACAGACCCAGTGTTCCTCAAGGACCACTGAACAGTGTATACACCT 1180	
DB	1890	ValGluProSerLysProThrValLeuArgSerThrSerThrSerProVal----- 1907	

Db 1181 TTGGATCCAACTGCTCAGCTACGAATCATGCTCACTGCTGCGGGCCCTTCAATAAGC 1240
Qy 1908 -----ArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeu 1925
Db 1241 CAAGCCCTGCCAGGCTCCGGTTACACACTGCTCGGATGCC-----CTGGCTGCTCTGTG 1297
Qy 1926 AspGlyValTyProThrLeuMetGluProValLeuLeuProLys-----Glu 1941
Db 1298 GATGCTGCAGCTTCTGCACCCAGATGATGTGTCACAAACAAAAGAGAGTAAGCATGAA 1357
Qy 1942 AlaProArgValAlaArgProGluArgProArgAlaLeu----- 1954
Db 1358 GCTCCAGGTTAGAGAAATTTGAGAGCAGGTCAGCAGCAGTGTAGTGAACAGCAGCAG 1417
Qy 1955 -----ThrGlyHis 1957
Db 1418 CTAGAGCAGAAACCCCTGGAGGTGGAGAGATGCTGTTCAAGTGTATTATACACTTCTTCA 1477
Qy 1958* AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSer----- 1974
Db 1478 GCCTTTCCAGTGGCAAGCCC-----CAGCCTCATTTCTCAGTAGTTAT 1522
Qy 1975 ProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIle 1994
Db 1523 TCTGAGGTGGGAAGATAAAGGCCCT-----CCTCCAAATCCAGATATAGGAAGAG 1576
Qy 1995 AlaArgThrProAlaLysAsn-----LeuAlaProHisHisAlaSerProAspPro 2012
Db 1577 CTAAGGACCAGAGGAGACTACCATTTACTGCTGCTCACTTCACTAGAGTGTATCATCACC 1636
Qy 2013 AlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSer 2032
Db 1637 CGGCAATTTGCTCGGACACAGATCGGAGGAACTGCTCTCAAGTTCAAGTCTTCT 1696
Qy 2033 IleGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSerTySerProGluGly 2052
Db 1697 -----AGTAGCTTATCTTCTCACAGGTATGAACACCTAGCGATGCT 1738
Qy 2053 ValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLys 2072
Db 1739 ATTGAGTGTAAGTCTCGCAGCTCACTCGGCCCCAGGAGAACTGCAGACCTAT 1798
Qy 2073 HisLeuGluLeuLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGly 2092
Db 1799 CAGCCAGAGGTGTTAAGGCAATCAAGCGGAATATGATCTTACCAGACATATGAAGGA 1858
Qy 2093 ProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSer 2112
Db 1859 CCA-----TTACATCACTATCGACCA-----CAG 1882
Qy 2113 GlnProSerSerProLeuLeuGlnThrAlaProGly----- 2125
Db 1883 CAGGAATCACCATCTCCCCAACACAGCTGCCCTTCTTCCAGGCGAGGGAATGGGG 1942
Qy 2126 ---ValLysGlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThr 2144
Db 1943 CAAGTGGCCAGGACCCATCGGCTGATCACACTTCTGATCACATCTGTCAAATTATCACA 2002
Qy 2145 GlnAspTyThrArgHisHis-----ProGlnGlnLeuSerAlaProLeu 2159
Db 2003 CAAGATTGTGTAAGAAATCAAGTTTCTCTCGCAGACTCCCCAGCAG-----CCTCCT 2053
Qy 2160 ProAlaProLeuTySerPheProGlyAla-----SerCysProValLeuAspLeuArg 2177
Db 2054 ACTTCTACATCCAGAACTCACTTCTGCTTTGTTGTTATCTACACCTGTG-----AGG 2104
Qy 2178 ArgProProSerAspLeuTyLeuProProAsp-----HisGlyAla 2192
Db 2105 ACTAAACATCAACCGTTACAGCCAGATCCAGGCTCAGTCTGTCCATCATCAAGA 2164
Qy 2193 Pro---AlaArgGlySerProHis-----SerGluGlyGlyLysArgSer 2206
||| :||| ||||| ||| ||| |||

RESULT 33
AF303585
LOCUS

AF303585 3997 bp mRNA linear PRI 01-FEB-2003

Db 2165 CCAGTTCAAGGGTCTCTCCAGAAATCTTGTGGACAAATCCAGGGGAAGTAGCCCTGGA 2224
Qy 2207 ProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSer 2226
Db 2225 AATCCCGAGAGAGAGTACGTC-----TCTTCGGAGCCCTACGAGCCCATCTCC 2275
Qy 2227 ProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyProLeuLeuTy 2246
Db 2276 CCACCCAG-----GTTCCGGTTGTGCATGAGAAACAGGACAGCTTCTCTTGTCT 2329
Qy 2247 ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer 2266
Db 2330 CAGAGGGCGCAGAGCCCTGCAGAGCAGAGAAATGATCCCGCTCACCAGGAGTAAAGC 2389
Qy 2267 GlnProProAlaPhePheSerLysLeuThrGlnSerAsnSerAlaMetValLysSerLys 2286
Db 2390 TACTTGCTTCATCTTCCACCAAGCTT---GAAATACATCACCCATGGTTAATCAAG 2446
Qy 2287 LysGlnGluLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyAsnIle 2306
Db 2447 AAGCAGGAGATTTTTCGTAAGTTGAACTCTCTGCTGGAGGTGACTCTGTATATGCGAGCT 2506
Qy 2307 SerGlnProGlyThrGluLeuPheAsnMetProAlaIleThrGlyThrGlyLeuMetThr 2326
Db 2507 GCTCAGCAGGAACTGAGATCTTTAATCTGCCAGCGTTACTACGTCAAGCTCAGTTAGC 2566
Qy 2327 TyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIle 2346
Db 2567 TCTAGAGCCATCTTTTGTGCTGATCTGCCAGT---AATCTTGGGCTGGAGACATATC 2623
Qy 2347 ArgLysAlaLeuMetGlyLysTyAspGlnTrpGlu-----Ser 2360
Db 2624 AGGAAGCTCTCTGGGAAGCTTTGATGACAAAGTTCAGGATCATGGAGTTGTCTATGCTCC 2683
Qy 2361 ProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAla 2380
Db 2684 CAGCCTATGGAGTA-----GTCCCTGGTACTGCCAACACCTCAGTT----- 2725
Qy 2381 MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly 2400
Db 2726 -----GTGACCAGTGTGAGACACGAGAGAGAGAGAGGAGCCCATCACCTCATTCAGGA 2779
Qy 2401 Gly-----LysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAla 2418
Db 2780 GGAGTTTGCAAAACCAAGCTGATCAGCAAGTCAACAGCAGCAAAATCTAAGTCTCTATA 2839
Qy 2419 Pro-----GlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGlu 2436
Db 2840 CCTGGCAGGCTACTTTAGGAACGGAACGGCCCTCTTTCAGTCTCTCTGTATCATTCAGAA 2899
Qy 2437 GlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSer 2456
Db 2900 GGGATTTACATAGCAGACGCCA-----GGTGGGCTGGGAGAGACAGAGCCCTCTTCA 2953
Qy 2457 AlaGlySerThrProPheProTyAsnProLeuIleMetArgLeuGlnAlaGlyValMet 2476
Db 2954 ACAGGCTCAACTCAGTTTCTTATAACCTCTGACTATGCGGATG-----CTCAGC 3004
Qy 2477 AlaSerProProProGlyLeuProAlaGlySerGly-----ProLeuAlaGlyProHis 2495
Db 3005 AGTACTCCACCAACACCGATTGTCATGTGCTCCCTCTCGGGTGAACCAAGCAGCTCTCAC 3064
Qy 2496 -----HisAlaTrpAspGluProLysProLeuLeuCysSerGlnTyGluThr 2512
Db 3065 CAACAGAACAGGATCTGGAGCGAGAGCCTGCCCACTGTCTCTCAGCACAGTACGAGACC 3124
Qy 2513 LeuSerAspSerGlu 2517
Db 3125 CTGTCGATAGTAGT 3139
||| :||| ||||| ||| ||| |||

[illegible]

Db 1515 AGATATGAGGAAGAGCTTAGCAGCAGAGGGAAGAGTACCATTACTGCGAGCTTAACCTTCATA 1574
Qy 2008 SerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLeuThrGln 2027
Db 1575 GACGTGATCATCACCCGCAAAATTCCTCGCACAAGGATGCGAGGAACGTGGCTCTCAA 1634
Qy 2028 SerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSer 2047
Db 1635 AGTTTCAGACTCTTCT-----AGTAGCTTATCTTCTTCACAGGTATGAA 1676
Qy 2048 TyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAsp 2067
Db 1677 ACACCTAGCGATGATATGAGGTATAGTCTCCAGCTCACCTCGCCACCCAGAG 1736
Qy 2068 LysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArg 2087
Db 1737 AAACCTGCAGACCTATCAGCCAGAGGTTCTTAAGCAAAATCAAGCGGAATGATCCTACC 1796
Qy 2088 ProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArg 2107
Db 1797 AGACAATATGAGGACCA-----TTACATCACTATCGA 1829
Qy 2108 ProLeuProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGly----- 2125
Db 1830 CCA-----CAGCAGGAATCACCATCTCCCAACAACAGCTGCCCTTCTTCACAG 1880
Qy 2126 -----ValLysGlyHisGlnArgValValThrLeuAlaGlnHisIle 2139
Db 1881 GCAGAGGAATGGGGCAAGTCCCGAGGACCATCGGCTGATCACACTTGCTGATCATC 1940
Qy 2140 SerGluValIleThrGlnAspTyHisArgHis-----ProGlnGln 2154
Db 1941 TGTCAAATATATCACACAAGATTTTGTCTAGAAATCAAGTTTCTCGCAGACTCCCCAGCAG 2000
Qy 2155 LeuSerAlaProLeuProAlaProLeuTySerPheProGlyAla-----SerCysPro 2172
Db 2001 -----CCTCTACTTCTACATTCAGAAATCCAGAACTCCTCTGCTTTGGTATCTACACCT 2051
Qy 2173 ValLeuAspLeuArgArgProProSerAspLeuTyLeuProProAsp----- 2189
Db 2052 GTG-----AGGACTAAACATCAACCGTTACAGCCCGAATCCAGGCTCAGTCT 2102
Qy 2190 -----HisGlyAlaPro-----AlaArgLysProHis-----SerGlu 2201
Db 2103 GTCCATCATCAAGACCAAGGTTCAGGCTCTCCAGAAATCTTGTGGACAAATCCAGG 2162
Qy 2202 GlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGly 2221
Db 2163 GGAAGTAGGCGCTGAAATCCCCAGAGAGGAGTCACTG-----TCTTCGGAGCCC 2213
Qy 2222 IleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVal 2241
Db 2214 TACGAGCCCATCTCCCAACCCAG-----GTTCCGGTTGTGCATGAGAAACAGGACAGC 2267
Qy 2242 TyrProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSer 2261
Db 2268 TTGTGCTCTGTCTCAGAGGGGGCGAGAGCTGCAGAGCAGAGGAATGATGCCCGCTCA 2327
Qy 2262 ProGlyAnthrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
Db 2328 CCAGGGAGTAAAGCTACTTGCCTTCACTTCTTCAACCAAGCTT-----GAAATATACATCACCC 2384
Qy 2282 MetValLysSerLysGlnGluLeuAsnLysLysLeuAnThrHisAsnArgAsnGlu 2301
Db 2385 ATGTTAAATCAAAAGACAGAGATTTTTCGTAAGTGAACCTCCTCTGGTGGAGGTGAC 2444
Qy 2302 ProGluTyHisLeuSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
Db 2445 TCTGATATGCGAGCTGCTCAGCCAGGAACCTGAGATCTTTAATCTGCCAGCAGTTACTACG 2504
Qy 2322 ThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
Db 2505 TCAGGCTCAGTTAGCTCTAGAGGCCATTCTTTTCTGATCCTGCCAGT---AATCTTGGG 2561

Qy 2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyArgProGlnTrpGluGlu----- 2359
Db 2562 CTGGAAGACATTAATCAGGAAGCTCTCATGGAAGCTTTGATGACAAAGTTTGAGGATCAT 2621
Qy 2360 -----SerProProLeuSerAlaAsnAlaPheAsnProProLeuAsnAlaSerAla 2375
Db 2622 GGAGTTGTCATGTCCTCCAGCCTATGGAGTA-----GTGCTGGTACTGCCAACACC 2672
Qy 2376 SerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
Db 2673 TCAGTT-----GTACCAGTGTGTGAGACACGAAGAGAGGAAGGGAGGCCA 2717
Qy 2396 SerProGlyGlyGly-----GlyLysAlaLysValSerGlyArgProSerSerArgLysAla 2414
Db 2718 TCACCTCATTCAGGAGTTTGCACCAACCAAGCTGATCAGCAAGTCAACAGCAGCAAAATCT 2777
Qy 2415 LysSerProAlaPro-----GlyLeuAlaSerGlyAspArgProProSerValSerSer 2432
Db 2778 AAGTCTCTTATACCTGGGCAAGCTACTTGAAGAACGGAACGGCCCTCTTCAGTCTCTCTCT 2837
Qy 2433 ValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAsp 2452
Db 2838 GTACATTCAGAAGGGGATTACCATAGGCAGACGCCA-----GGGTGGGCTCGGAAGAC 2891
Qy 2453 ArgProSerSerAlaGlySerThrProPheProTyArgAsnProLeuIleMetArgLeuGln 2472
Db 2892 AGGCCCTTTCACAGAGCTCAACTCAGTTTCTTATAACCTCTGACTATGCGATG--- 2948
Qy 2473 AlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGly---ProLeu 2491
Db 2949 -----CTCAGCAGTACTCCACCAACCAACCGATTGCATGTGCTCCTCTGCGGTGAACCAA 3002
Qy 2492 AlaGlyProHis-----HisAlaTrpAspGluProLysProLeuLeuCysSer 2508
Db 3003 GCAGTCTCTCACCACAGAACAGGATCTGGAGCGAGAGCCTGCCCACTGCTCTCAGCA 3062
Qy 2509 GlnTyThrLeuThrLeuSerAspSerGlu 2517
Db 3063 CAGTACGAGACCTGTGCGATAGTGAT 3089

RESULT 34

AL590153/c

LOCUS

DEFINITION

133947 bp DNA linear VRT 29-JUL-2003
Zebrafish DNA sequence from clone RP71-71M17 in linkage group 8
Contains a novel gene and part of a novel gene for a protein
similar to mouse silencing mediator of retinoic acid and thyroid
hormone receptor (SMRT), complete sequence.

ACCESSION

AL590153.9

VERSION

GI:16973912

KEYWORDS

HTG; retinoic acid receptor; silencing mediator; SMRT; thyroid

SOURCE

hormone receptor.

Danio rerio (zebrafish)

ORGANISM

Danio rerio

REFERENCE

Laird, G.

Direct Submission

Submitted (20-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Nov 17, 2001 this sequence version replaced gi:15022264.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep/RF71-71M17> is
from a Zebrafish mixed sex BAC library VECTOR: pFAR2AC2
This sequence is the entire insert of clone RF71-71M17 This clone
was isolated from BAC library RPCI-71 created by Pieter deJong and
provided by Gareth Howell (Wellcome Trust Sanger Institute,
Cambridge, UK)
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
Clone-derived Zebrafish pUC subclones occasionally display
inconsistency over the length of mononucleotide A/T runs and
conserved TA repeats. Where this is found the longest good quality
representation will be submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat
discovery system (Zhirong Bao and Sean Eddy, submitted), and those
beginning 'dr' were identified by Rick Waterman (Stephen Johnson
lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
----- Genome Center

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk

FEATURES
source

Location/Qualifiers
1. 133947

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="RF71-71M17"
/clone.lib="RPCI-71"
complement(245..378)

/note="DR20 repeat: matches 109. .242 of consensus"
repeat_region
complement(380..481)
/note="Dr000409 repeat: matches 1. .102 of consensus"
repeat_region
499. .872
/note="HE1_DR1 repeat: matches 2. .384 of consensus"
repeat_region
873. .882
/note="2.5 copies 4 mer GAA 20% conserved"
repeat_region
1035. .1046
/note="6.0 copies 2 mer TG 24% conserved"
repeat_region
1049. .1065
/note="8.5 copies 2 mer AG 34% conserved"
repeat_region
1194. .1204
/note="2.8 copies 4 mer GTT 22% conserved"
repeat_region
1287. .1301
/note="15.0 copies 1 mer T 30% conserved"
repeat_region
1393. .1405
/note="2.2 copies 6 mer TCTCAT 26% conserved"
repeat_region
1404. .1420
/note="2.1 copies 8 mer TTGATGAT 34% conserved"
repeat_region
1448. .1457
/note="10.0 copies 1 mer A 20% conserved"
repeat_region
complement(1688..2225)
/note="Dr000393 repeat: matches 3. .549 of consensus"
repeat_region
2252. .2412
/note="dr7759 repeat: matches 2. .131 of consensus"
repeat_region
2755. .2767
/note="2.6 copies 5 mer TTGA 26% conserved"
repeat_region
complement(2841..3117)
/note="dr1203 repeat: matches 40. .310 of consensus"
repeat_region
3141. .4315
/note="Dr000241 repeat: matches 3. .1153 of consensus"
repeat_region
complement(4343..4836)
/note="dr994 repeat: matches 2. .493 of consensus"
repeat_region
4912. .4925
/note="14.0 copies 1 mer A 28% conserved"

repeat_region
5050. .5060
/note="2.2 copies 5 mer TGA 22% conserved"
repeat_region
complement(5528..5702)
/note="Dr000365 repeat: matches 5. .182 of consensus"
repeat_region
6032. .6046
/note="7.5 copies 2 mer AC 30% conserved"
repeat_region
6351. .6360
/note="2.0 copies 5 mer TTA 20% conserved"
repeat_region
6430. .6447
/note="3.0 copies 6 mer TAAAA 36% conserved"
repeat_region
complement(6463..6739)
/note="Dr000085 repeat: matches 217. .518 of consensus"
repeat_region
complement(6736..6845)
/note="dr1076 repeat: matches 1. .118 of consensus"
repeat_region
complement(6852..7008)
/note="Dr000325 repeat: matches 89. .248 of consensus"
repeat_region
complement(7038..7356)
/note="Dr000208 repeat: matches 441. .745 of consensus"
repeat_region
7362. .7385
/note="2.2 copies 11 mer TTTAAAGTCA 39% conserved"
repeat_region
7420. .7440
/note="2.1 copies 10 mer TTGTACATG 33% conserved"
repeat_region
7457. .7474
/note="2.6 copies 7 mer AATATT 27% conserved"
repeat_region
7468. .7481
/note="2.8 copies 5 mer ATTTA 28% conserved"
repeat_region
7669. .7678
/note="5.0 copies 2 mer AG 20% conserved"
repeat_region
7772. .7782
/note="5.5 copies 2 mer TG 22% conserved"
repeat_region
7899. .7918
/note="3.3 copies 6 mer CTTTCT 22% conserved"
repeat_region
7904. .7918
/note="7.5 copies 2 mer CT 21% conserved"
repeat_region
7944. .7955
/note="2.0 copies 6 mer CACAGT 24% conserved"
repeat_region
8150. .8169
/note="2.0 copies 10 mer TCATGATTA 40% conserved"
repeat_region
8349. .8363
/note="2.5 copies 6 mer TTGTAT 21% conserved"
repeat_region
complement(8394..8560)
/note="HE1_DR1 repeat: matches 206. .374 of consensus"
repeat_region
complement(8530..8606)
/note="Dr000009 repeat: matches 771. .848 of consensus"
repeat_region
complement(8549..8614)
/note="DNA9TA1_DR repeat: matches 1196. .1261 of consensus"
repeat_region
complement(8629..8681)
/note="HE1_DR1 repeat: matches 334. .384 of consensus"
repeat_region
8682. .8704
/note="3.3 copies 7 mer TATTATT 37% conserved"
repeat_region
8684. .8704
/note="5.2 copies 4 mer TTAT 35% conserved"
repeat_region
8710. .8721
/note="Dr000377 repeat: matches 153. .231 of consensus"
repeat_region
complement(8722..8781)
/note="HE1_DR1 repeat: matches 47. .115 of consensus"
repeat_region
8876. .8885
/note="2.0 copies 5 mer ATATA 20% conserved"
repeat_region
8951. .8990
/note="2.9 copies 14 mer AATTATGATCAATC 46% conserved"
repeat_region
9147. .9156
/note="2.5 copies 4 mer GTGT 20% conserved"
repeat_region
complement(9327..9377)
/note="LOOPERN2_DR repeat: matches 845. .895 of consensus"
repeat_region
9622. .9642
/note="10.5 copies 2 mer CA 24% conserved"
repeat_region
9794. .9805
/note="6.0 copies 2 mer TG 24% conserved"
repeat_region
9820. .9829
/note="5.0 copies 2 mer GA 20% conserved"
repeat_region
10116. .10565
/note="Dr000322 repeat: matches 5. .519 of consensus"
repeat_region
10676. .10806


```
Db 76289 CCGTGGCGGGACAGCGCAGCAGCGCG-TCAAGTAGGAGAAAGACTCAAAACAAAGTGA 76231
Qy 2031 heSerileGlnGluLeuGlySerLeuGlyTyrHisGlySerTyrSerProG 2051
Db 76230 TGTCATACAGAACAGACTTCGAGACTCGGT-----AAGACCAACATGA 76183
Qy 2051 luGlyValGluProValSerProValSerProSerLeuThrHisAspLysGlyLeuP 2071
Db 76182 CAGCGGCCAACCTTCATAAACCGGATAATCATGTCATCAAAATTCCTGTGTATCGCGGATGC 76123
Qy 2071 roLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeu-ArgProLysGln 2090
Db 76122 CAGAG-----ACTGGTGGCGCTCGCGACCAACGCGCA 76093
Qy 2091 ProGlyProValLysLeuGlyGluAlaAlaHisLeuProHisLeuArgProLeuPro 2110
Db 76092 CCTGTGTGTGTAAGATGCTCTGGGACCTTGGGCGGTTA----- 76054
Qy 2111 GluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGln 2130
Db 76053 -----ACGCCACATCAAGCCCTCCCTGCCCATTCCTCTCTTCTACCCAGTCCACC 76000
Qy 2131 ArgValValThrLeuAlaGlnHisLeuSerGluValLeuThrGlnAspTyrThrArgHis 2150
Db 75999 AAAACCCCTCTCCCTCCCTCTCCCTC----- 75973
Qy 2151 HisProGlnGlnLeuSerAlaProLeu-----ProAlaProLeuTyrSerPhe 2166
Db 75972 ---CCTCCTCCCTCATCATTCCTCTGCCCATCTCGCCCATCTCGCCCTTACGCTGTGG 75916
Qy 2167 ProGlyAlaSerCysProValLeuAspLeuArg 2178
Db 75915 CCTAGTCACGCTGTGTGTGTGTGTCATTAATGAGA 75880

RESULT 35
AX677743
LOCUS AX677743 650 bp DNA linear PAT 27-MAR-2003
DEFINITION Sequence 521 from Patent WO02086122.
ACCESSION AX677743
VERSION AX677743.1 GI:29335148
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Legrain, P. and Daviet, L.
TITLE Protein-protein interactions in adipocytes
JOURNAL Patent: WO 02086122-A 521 31-OCT-2002;
Hybridgenics (FR)
FEATURES
source
location/Qualifiers
1..650
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 1.57e-15 Length: 650
Score: 1113.00 Matches: 216
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.42% Indels: 0
DB: 6 Gaps: 0

US-09-522-753-5 (1-2517) x AX677743 (1-650)
Qy 203 GluAlaAlaLysProGluProGluLysProValSerProProLysGluSerLys 222
Db 2 GAGCTCCCAAGCCCGCCGAGCTTGAAGCCCGGTGTCACCGCCCGCCATCGAGTCAAG 61
Qy 223 HisArgSerLeuValGlnIleLeuTyrAspGluAsnArgLysLysAlaGluAlaHis 242
```

```
Db 62 CACCGCAGCCTGTGTGCAGATCATCTACGACGAGAACCGAAGAGGTGAAGTCCACAT 121
Qy 243 ArgIleLeuGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAsp 262
Db 122 CGGATTCTGAAGCCCTGGGGCCCGAGGTGAGCTGCGCTGTATCAACACCCCTCCGAC 181
Qy 263 ThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuLeu 282
Db 182 ACCCGGCAGTATCATGAGAACATCAANAATAAACAGCGCATCGGAAGAACTAATCTTG 241
Qy 283 TyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyr 302
Db 242 TACTTCAAGAGGAGGAATCAGCTCGGAACAATAGGAAGCAGAAAGTTCTGCAGCGTAT 301
Qy 303 AspGlnLeuMetGluAlaLeuLysValGluArgIleGluAsnAsnProArgArg 322
Db 302 GACGAGCTCATGGAGGCTTTGGAAAAAAGGTGAGCGCATCGAAAAACAACCCGCGCGG 361
Qy 323 ArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLys 342
Db 362 CGGSCCAAGGAGAGAGAGGTGGCGAGTACTACGAAAGCAGTTCCCTGAGATCCGCAAG 421
Qy 343 GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer 362
Db 422 CAGCGCAGCTGCAGGAGCGCATGCAGCAGGAGGTGGCGCGCAGCGCGGCTGTGCC 481
Qy 363 MetSerAlaAlaArgSerGluHisGluValSerGluIleLeuAspGlyLeuSerGluGln 382
Db 482 ATGTGCGCGCCCGCCGACGAGCAGCAGGTGTTCAGAGATCATCGATGGCTCTCAGAGCAG 541
Qy 383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402
Db 542 GAGAACCTGGAGAAGCAGATGGCGCAGCTGGCGCTGATCCGCGCCATCTGTACGACGCT 601
Qy 403 AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAsp 418
Db 602 GACCAGCAGCGCATCAAGTTTCATCAACATGAACGCGCTTATGGCCGAC 649

RESULT 36
AX6775223
LOCUS AX6775223 11296 bp mRNA linear INV 06-DEC-1999
DEFINITION Drosophila melanogaster SANT domain protein SMRTER (Smrter) mRNA,
complete cds.
ACCESSION AF175223
VERSION AF175223.1 GI:5815244
KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
REFERENCE 1
AUTHORS Tsai, C.-C., Kao, H.-Y., Yao, T.-P., McKeown, M. and Evans, R.M.
TITLE SMRTER, a Drosophila nuclear receptor coregulator, reveals that
JOURNAL ECR-mediated repression is critical for development
Mol. Cell 4 (2), 175-186 (1999)
MEDLINE 99417957
PUBMED 10488333
REFERENCE 2 (bases 1 to 11296)
AUTHORS Tsai, C.-C. and Evans, R.M.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Gene Expression Lab, The Salk Institute,
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
location/Qualifiers
1..11296
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
1..11296
/gene="Smrter"
283..10692
/gene="Smrter"
```


Db 2479 CGAAATTTGAAGCGGCAACGCTGGCGGCACCAATCAGGGAGCTGGTGAGAAAGTACACCAAG 2538
Qy 305 LeuMetGluAlaLeuGluLysValGluArgileGluAsnAenProArgArgAla 324
Db 2539 GATCAGCGGATTTGGCAGCAGATGCGAGCGTATGGAGCGCAGTGGCAAGAGAAAGCA 2598
Qy 325 LysGluSerLysValArgGluTyrGluLysGlnPheProGluLeuLeuArgLysGlnArg 344
Db 2599 CGCAGGCAAAATCCGGAATTTTCGAAAGGTTTACGGAGCTAAGGAAACAGCGC 2658
Qy 345 GluLeuGluArgMetGlnSerArgValGluGlnArgGlySerGlyLeuSerMetSer 364
Db 2659 GAGATAGGAGACGTTTC---AATCGCGTCCGTTCCGCG--- 2694
Qy 365 AlaAlaArgSerGluHisGluValSerGluLeuLeuLeuLeuLeuLeuLeuLeu 384
Db 2695 ---ATTAAATCCGAAGCGGATCTCGAGGAGATCATGATGCGCTTCGAGGAGCGCGTTG 2751
Qy 385 LeuGluLysGlnMetArgGlnLeuAlaValleProProMetLeuTyrAspAlaAspGln 404
Db 2752 GAGGATAAAAGATGCGATCTCATCGCGTTATACCGCTCTGTATGACGATGCGCGCAG 2811
Qy 405 GlnArgileLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys 424
Db 2812 AGCGGATCGCGCTCATATGAAATGGAATTCGAGGATATGCGGAGTATGCGGAGTTTCATCA 2871
Qy 425 AspArgGlnValMetAsnMetTyrSerGluGlnLysGluThrPheArgGluLysPhe 444
Db 2872 CAGCGTAAGGCCCTCAACATGTGACCGCGCGGAGAGGAGACCTTTAAGGAGAGTAT 2931
Qy 445 MetGlnHisProLysAsnPheGlyLeuLeuAlaSerPheLeuGluArgLysThrValAla 464
Db 2932 TTGCAGCATCCGAAAAAATTTGAGGACCATTCGCGCTAGTTTGGATCGCAATCGCCACAG 2991
Qy 465 GluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArg 484
Db 2992 GATTGCGTTCGCTACTACTACTCAGCAAGAAAGACCGAGAAATACAGACGCTGCTCGA 3051
Qy 485 ArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 503
Db 3052 AAATCCCGCAACGCGCGCGAGTTCGCGTAATCCAGCCAGGCTCAGCGCGCACACCGC 3111
Qy 504 ----- 3171
Db 3112 CAGTGCAATTATCGATGACCACTGGCGTTATGACGGGATTCAGCGCGCAACGCA 3171
Qy 509 GlnGlnProMetProArgSerSer-----GlnGluLysAspGluLysGluLys 525
Db 3172 CAAAGTCGCGCGTCTGTTTCATCCCGCTGCTGAGCGTGAGCGTGGCGAACGTCGCGCC 3231
Qy 526 GluLysGlu----- 532
Db 3232 GAGCGTGAGCGTGTGSCAGAGAAAGCGCAGCTGATGCTGCCAAAGCGGCTGAAAGTGC 3291
Qy 533 GluLysProGlu----- 537
Db 3292 GCCGAAAGCAAGCGCGCGCACTAAGGCTGTGCAAGCAACGCGTGGCGTGAAGAGTG 3351
Qy 537 ----- 537
Db 3352 GCCAAAGCGGCTGCAGCAGCGCGCGGCGGAGCACTACCGCCCAACCGCAACCA 3411
Qy 537 ----- 537
Db 3412 ACCACATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCA 3471
Qy 538 -----ValGluAsnAspLysGluAspLeuLeu 546
Db 3472 TCATCGAGCGCAAGTCCAGCGACATTCGCGAGAAATAGCGCAGATAAAACGATGCGCGC 3531
Qy 547 Lys-----GluLysThrAspAspThrSerGly----- 555
Db 3532 AAAACGGCTAGCGCGGATAGAAATGTCAGCAACAGCTGCTGGAGCAACCGCAACCGGGA 3591

Qy 555 ----- 555
Db 3592 ACTCCGACAGCAGCAACTACACCTGCAACAGCAACAGCACCACAGAGATCTCAGCGGCG 3651
Qy 556 -----GluAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThr 571
Db 3652 GCGAGGCCAAGAGCAAAATGCGAGGAGGAGCAGCTGCCACAGCGCGGAGCAGCAACG 3711
Qy 572 AlaAsnSerGlnGlyArgArgLysGlyArgileThrArgSerMetAlaAsnGluAlaAsn 591
Db 3712 GTGGCCACTGCGGTACACCGGCAACCGGACCAAGTGGCGCCAGTGTGGCGAGCGGCA 3771
Qy 592 Ser-----GluGluAla 595
Db 3772 ACGCCAACCGCGGCCACCGGCGCAGCTGCGCCCAAGGGTGTGGCAAAACCGAAACAGCG 3831
Qy 596 IleThrProGlnGlnSer----- 601
Db 3832 ACTGAGCGCAGCGGGAACGGCTGCTAAGGGAGCTGACTCTCGGCCAGATGCCAACAGTCCG 3891
Qy 602 ---AlaGluLeuAlaSerMetGluLeuAsn-----GluSerSerArgTyrThr 616
Db 3892 CTGGCCAAAATCTGCTTCGAAGGCTATCAATGCCAGGGCTATATATGCTAT---CGGTGGCA 3950
Qy 617 GluGluGluMetGluThrAlaLysLysGlyLeuLeu-----GluHisGlyArgAsnTrp 634
Db 3951 CAGCAGCAGCAGCAGCAGCAATGCAACGGCGCGCTCTGCTCCAGTGCAGGCGGTACCTT 4010
Qy 635 SerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsn----- 651
Db 4011 GAATGGCTTTAAGCCTGGCTATCAAAACCGTGGTTCATGCCCAATGTAAGGATCCACGG 4070
Qy 652 -----PheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluLeuGln 668
Db 4071 TGGCGATGATCTGGTG---CGAACGCTGGTGAGCGCGCCGATCGTTGACGCAACGA 4129
Qy 669 ---HisLysLeuLysMetGluLysGluArgAsnAlaArgArgLys---LysLysLysAlaP 687
Db 4130 ATGCATCGATTGGCAGTGCGGATAAATCGTGAAGACACACGCCATCATCCAGGCGCG 4189
Qy 687 roAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluGluMetGlu 707
Db 4190 CAAATCTCATCTTCAACCGCGCG---AATGAGAGTAGTAGTG 4231
Qy 707 LysArgLysVal-----SerGlyAsn----- 713
Db 4232 GTGCGGAGTGAATACCTATGCGCCACACCGCAACCGCGGCGCAATTCCTTGGACAGA 4291
Qy 714 -----GluGluGluMetValGluGluAlaGluAlaLeuAlaLeuHisAlaSerGly 729
Db 4292 AGCTGAAGGCTGCCCAAGTTGAGGTTTGGAGCTGGCAATGAATTCGATTCGGATGTTA 4351
Qy 729 snGluValProArg-----GlyGluCysSerGlyProAlaThr- 741
Db 4352 GTGAATCCAGAGGAAACGATTCGAATTCGATTCGCTGAGGCGAGGTTGAAATGCAACGT 4411
Qy 742 -----ValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAla 760
Db 4412 CAGCTATGACCAATAGCAGCAGCCAGTGGCAGCATGAACATCAGCAATAGTACCGGATGA 4471
Qy 760 YsAspThrGlyGlnAsnGly----- 766
Db 4472 AGGCCAATGCCAAGGATGATCGATGATGCGCAAAACATCGATGCGCATCCACATCGTCG 4531
Qy 767 -----ProLysProAlaThr----- 772
Db 4532 CATCGGTTGTGTACATCTACGCGCATCGGCATCTCATCTGTCATTCATTCGCGTCGT 4591
Qy 772 ----- 772
Db 4592 CCATGCTGTTGATAAGTGCAGCATCCGTAATGTGCCAGCGGCGGTGCAACGTCCTCGT 4651

QY 773 -----LeuGlyAlaAspGlyP 778
Db 4652 CCACGGCGCAACACACGGCCACAGCTTCGGCGATCTCGTTCGGCTATTGGCGGATGCA 4711
QY 778 roProProGly----- 781
Db 4712 GTGGCAACTCGATGTCATGCAATGCAACAGAGATTCTCCCTCGATGCGCAAGACAATGG 4771
QY 782 -----ProProThrProArgAThrSerA 791
Db 4772 CCAGTCTTTGTGTGCAAGCGGAGCGGTCTCCCGCACCCGTCCACTAAAGAGGGGC 4831
QY 791 rGAla-----ProileluProThrProLaser----- 800
Db 4832 GTGGCCAGCAGTAGTGGCATTCGGACGACAAACATACCGCGGGTCCCGCTCTCGAATA 4891
QY 801 -----GluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProP 819
Db 4892 GCTGTCAATCAATCGTGTGCGAGCGCTATCCAAATTGTCGCTACCCACATGCCCAA 4951
QY 819 roProVal-----ValProLysGluGluLysGluGluGlu 831
Db 4952 ATCCCAAGGATCGAGCCAGGCTCGGGAACATTCTCCCTCCGTCTCTCGAA----- 5004
QY 831 hrAlaAlaAlaProProValGlu-----GlyGluGluGluLysProProAlaA 848
Db 5005 -----CTGGCCCGCAGGTGGTGTGATCGGTGATGCGCGGAATTTCAAAATACCCCGCATG 5059
QY 848 la----- 848
Db 5060 CAACACGCTGCTGTTCGGCGCTGCCTATGCGCATTCGTCGACAGCTCATCCGAGCTGA 5119
QY 849 -----GluGluLeuAlaV 853
Db 5120 ATCTCACGATGGTCCAGTGGCGTCCGCGCAGCGAGTGGCGGCGATGACGCGATG 5179
QY 853 alAspThrGlyLysAlaGlu-----GluProValLysSerGluCysThrGluGluAlaG 871
Db 5180 TCAGTACCTCGTGTGTGATGAACAGAGCGCGGCGGATCCGATACCGTTCCTGCA-G 5238
QY 871 luGluGly-----ProAlaLysGlyLysAspAlaGluAlaAlaG 884
Db 5239 ATCCGGAGAAATTCACGGCGCACAAAGTCCCTTACGATGGTCAACAGCAACAGCAACAG 5298
QY 884 luAlaThrAlaGluGlyAlaLeuLysAla-GluLysLysGlu----- 897
Db 5299 CAGCAACAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAACAGCAACTATCG 5358
QY 897 ----- 897
Db 5359 CAGCCACAGCCACCTCCGCGCAGCTCCGCGAGCAGAGAGGGTTCGTTCGGACGAGTGGT 5418
QY 898 -----GlyGlySerGly 901
Db 5419 GATCAGGGTACCACTGATTATACACCAACCCGAATGACGAGCAAGTCCGGATCGGGC 5478
QY 902 ArgAlaThrAla----- 906
Db 5479 GGTTCGCAAAACAGCGGCGCAATGAACGCCTATGCGCACCCCGCGTGGCCAGCCACC 5538
QY 907 LysSerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAlaAspGluVal 926
Db 5539 AAGAACAAAGCAGCAGGAGGAAATACGATTCTTCGGCCACGAAACGCGGATCAGGAG 5598
QY 927 AspGlu----- 928
Db 5599 AACGAAATCTCGCGGCCAATCGTCAGAGTCCCAAGGTACTATTCCATGGCCATGGACAT 5658
QY 929 AlaGluGlyGlyAspLysSerArgLeuLeuSerProArgPro-----SerLeuLeu 945
Db 5659 GGCATGGTGGTTCATGCCAATATGTGGTGTGCTGCACCGCGCGTGCCTATATGGG 5718
QY 946 ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLys-----ProLeuAsp 962

Db 5719 ACGGTGGTGGTGTTCACACGAGTGGAGCTGTGGGCAACAGGTCAACGACCGATTAGT 5778
QY 963 LeuLysGlnLeuLysGlnArgAlaAlaAlaAlaProProleGln-----ValThrLys 980
Db 5779 ATG-----CGACGCGAGCGGTCAACAAATGTTCCAGGATTCGCTATTCTCG 5823
QY 981 ValHisGluProProArgGluAspAlaAlaProThrLysProAlaProProAlaProPro 1000
Db 5824 GTAATTGAG-----CGTTCGCTGAAGCACAGGTCGCGCAACCGAAG 5865
QY 1001 ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArg 1020
Db 5866 GCGCAGCAGGTCAGCAGCAGGG-TCAGGGGCAAGGACAGGGCCAGGTCAGGTCAGAC 5924
QY 1021 GlyLysSerArgSer-----ProAlaProProAlaAspLysGluAlaPheAlaAlaGluAla 1039
Db 5925 TCCAGGTCAAAGCCAGTCCCATCCAGCAGCAACAGCAGCAGCAGCAGCAGCAATCTGC 5984
QY 1040 GlnLysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProValPro-Pr 1059
Db 5985 CAATAATCTGGAACCGAAGGAGCTTACCATTTGTAAGGGAATACCG-----CCAGGA 6035
QY 1059 oArgGluValLysAlaSerProHis-----AlaProAspProSe 1073
Db 6036 TCCGGGTATCTGAAGCAGCAGCAGCAGCAGCTGCGGAGCAGCAGCCGCCACCTC 6095
QY 1073 rAlaPheSerTyAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaAr 1093
Db 6096 GGTCCCGGTAGTTTGGCGCATGGC-----ACCTCCGTACA 6131
QY 1093 gProValLeuProArgProProThrIleSerAsnProProProLeuLysSerSerAlaLys 1113
Db 6132 GAAGCTAACACCCCGCCAGCTGCAGTGCACACCCAGCCACCG-----GC 6176
QY 1113 shisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLe 1133
Db 6177 TCATCCA-----CTGACACCCACGAGCATTTGGATGTGCGGGCAGCAACACCGAACCGCA 6233
QY 1133 uHisValProTySerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuPr 1153
Db 6234 TAGCTTAGCCACTTATCGTGTGTCATTCGCATGGGATGTTGGGAATTTGGTCATCC 6293
QY 1153 oLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluLeuSe 1173
Db 6294 GGGCCGATGGCCACCAAGTTCCCGGAGGAATCGGTGTGGAGACAGGCGACCATAC 6353
QY 1173 rPro-----ArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGl 1191
Db 6354 GCGGTTGTCAAGAGCTCTAGTGGATCATCGAAGCGCGCGCGCGGCGTCCAGTCA 6413
QY 1191 uAlaSerValLeu---ArgGlyThrAlaLeuGlySerValProGlySerIleThrly 1210
Db 6414 CTGACAGCCACACCCCGGAGACCAATAATCTACAAATGTCCAGTGGCGCATCCGCAACG 6473
QY 1210 sGlyIleProSerThrArgValProSerAspSerAlaIle-----ThrTyArgGlySe 1228
Db 6474 TGTATATACCA-----CCGTCGACGACTCGGTATACATCCCGGCCATCCGTCGCA 6524
QY 1228 rIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrIleThrArgIle---Il 1247
Db 6525 TACACAGCCCGGCCCATCCGAGCATTCGTGCGATGTCGATGTCAGCATACGCGAGCTGCAAGT 6584
QY 1247 eGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHi 1267
Db 6585 GCCGGAACCGGAGCGCAAAATTTGACATCAAGAACCGCGCGCGCGGATGTCA 6644
QY 1267 sValIleTyGluGlyLysLysGlyHisValLeuSerTyGluGlyGlyMetSerValTh 1287
Db 6645 CTCGCCGACACCGGA-----GCCGAGGATCTAGCTC 6677
QY 1287 rGlnCysSerLysGluAspGlyArgSerSerSer-----GlyProProHisGl 1303

Qy	1836	GlySerSerGlySerSerGlyGlyGlyGlySerSerArgProAlaSerHisSer	1855
Db	8615	---AAGAGCGGAGCAGATAGTCAGCGCATGGAGCGCATGTCGCGCAATGTGGTCA	8671
Qy	1856	HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgPro	1875
Db	8672	---CTGGATCAGGTGGCGCGCGGGGGAGGAC---	8710
Qy	1876	---SerValLeuHisAsnThrGlyMetIysGlyIle	1886
Db	8711	GTCAAAATTCTCGCGGCTCGGTTCCGAAACATGGTCCACCAAGCTCAA---	8758
Qy	1887	IleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSer---	1905
Db	8759	---TCCAGATCGGAAACGGGAATCGTACTATCGTCAGGCACACACGGCGGTCCAGTCCAG	8815
Qy	1906	ProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeu	1925
Db	8816	AGGATACGCCCGGCAACTGAGCGCCAGAGTCTGATCGATGCATTAT---TCAAGCATG	8872
Qy	1926	AspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgVal	1945
Db	8873	AGATCAATCGTTCCA-----ATGATGCCACCCTGCG-CCGGGTGCG-----	8913
Qy	1946	AlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProAla	1965
Db	8914	---GAAATTCCGCGCCCATCGTTCTGTTTCATGCTCCGTTG-----CGGCCACGT	8958
Qy	1966	ArgSer-----GlyLeuGluProAlaSerSerProSerLysGlySerGluPro	1981
Db	8959	GGTTCGGATCGGAGCGCGCAGCGGTACACGATCTCTACCGGCCAATGTGCTCCATCG	9018
Qy	1982	---ArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrPro	1998
Db	9019	ATGTACCTGCGCGATCTACGCCACCGCTCGATGGCGGAGCGGATCAATGCTTACGGCG	9078
Qy	1999	AlaLysAsnLeuAlaProHisHisAlaSerProAspProAla-----	2013
Db	9079	GAGAACATGGCAAGCCC-----AGTCTCGGGATCGCCAGTGTGATTAATTGAT	9132
Qy	2014	---ProProAlaSerAlaSerAspProHisArgGlyLysThrGln	2027
Db	9133	TTGGATCAGGAGCGCATCTCAGCGCGCAGCGCAGCAGTTCGCCAACAGCAGCAGCAGCAG	9192
Qy	2028	SerLysProPheSerIleGlnGluLeuLeuArgSerLeuGlyTyrHisGlySerSer	2047
Db	9193	CAAGTCCGCTCGTCCGTCATCTCGAGCTTAGATCCGTG-----CACGGTCAGCTA	9246
Qy	2048	TyrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAsp	2067
Db	9247	AGGACACCGACCTCG---CAATCTGGCGGATCAGCGCCAGTCCACAGCAGATTCATACG	9303
Qy	2068	LysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArg	2087
Db	9304	AAGACATTACG-----TTTGGGGAACTAACCGATTTCGATAATTACCGTACTACGGC	9357
Qy	2088	ProLysGlnProGlyProValLysLeuGlyGluAlaAlaHisLeuProHisLeuArg	2107
Db	9358	ACCAAT-----CCTCATCTGCGG	9375
Qy	2108	Pro-----LeuProGluSerGlnProSerSerPro-----	2118
Db	9376	CCACCATATGGCCCTATTTCAGGAGACGCGAGTCAATTCTGCGCGCGGATCGTGGAG	9435
Qy	2119	-----LeuLeuGlnThrAlaProGlyValLysGlyHis-----Gln	2130
Db	9436	CAGAACCGTCGTATGCAACAGAGGCCGAGGAAGCAACGACCAATCCACAGCAGCAGCAG	9495
Qy	2131	ArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHis	2150
Db	9496	CAGCAGCAACACACGACGACATCATCCGCCGAGCAGCAGCAGCAGCAGCAGCAACAC	9555
Qy	2151	HisProGlnGlnLeuSerAlaProLeuProAlaPro---LeuTyrSerPheProGly---	2168

[illegible]

Db 10485 A-----GGAAATTGCGAGCGGA-----CCAGGTGTGTCTC 10514
Qy 2431 rSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGl 2451
Db 10515 A-----10515
Qy 2451 uAspArgProSerSerAlaGlySerThrProPheProTyAsnProLeuLeuMetArgLe 2471
Db 10515 -----10515
Qy 2471 uGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLe 2491
Db 10516 -----GTCGGTCTC-----GGTGATCCGGCGCGCGGACCAGG 10550
Qy 2491 uAla-----GlyProHisHis-----2496
Db 10551 AAGCGGTGGCGCGCGGTGGCGGTGCACATTCAGTTCGCAAGCATCTGTCGGGT 10610
Qy 2497 -----AlaTrpAspGluGluProGlyProLeuLeuProLeuLeuCysSerGlnTyrgl 2511
Db 10611 TGCCTGTCGGCTTTCAGCGCGCGCGCGTGCAGTGCAGCGCGTGTGTGTCTCAAGTACGA 10670
Qy 2511 uThrLeuSerAspSerGlu 2517
Db 10671 TGCCTCAGCGCGCAAGAT 10689
RESULT 37
LOCUS BD221548 872 bp DNA linear PAT 17-JUL-2003
DEFINITION Human gene and gene expression product v.
ACCESSION BD221548
VERSION BD221548.1 GI:33031318
KEYWORDS JP 2002534055-A/2661.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 872)
Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.S.,
Reinhard,C., Giese,K., Randazzo,F., Kennedy,G.C., Pot,D.,
Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
Drmanac,S., Labat,L., Leshkowitz,D., Kita,D., Garcia,V., Jones,L.W.
and Crain,B.S.
Human gene and gene expression product v
Patent: JP 2002534055-A 2661 15-OCT-2002;
CHIRON CORP.HYSEQ INC
OS Homo sapiens (human)
PN JP 2002534055-A/2661
PD 15-OCT-2002
PF 13-MAY-1999 JP 2000548466
PR 14-MAY-1998 US 60/085426,15-MAY-1998 US 60/085537 PR
15-MAY-1998 US 60/085696,21-OCT-1998 US 60/105234 PR
27-OCT-1998 US 60/105877
PI LOUIS T WILLIAMS,JAIME ESCOBEDO,MICHAEL A INNIS,PABLO PI
DOMINGUEZ GARCIA,
PI JULIE SUDDUTH KLINGER,CHRISTOPH REINHARD,KLAUSE GIESE,FILIPPO
PI RANDAZZO,
PI GIULIA C KENNEDY,DAVID POT,ALTAF KASSAM,GEORGE LAWSON,RADOJE
PI DRMANAC,
PI RADONIR CRKVENJAKOV,MARK DICKSON,SNEZANA DRMANAC,IVAN LABAT,
PI DENA LESHKOWITZ,DAVID KITA,VERONICA GARCIA,LEE WILLIAM JONES,
PI BIRGIT STACHE CRAIN
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC
C12N5/00,C12O1/68,
PC C12N15/00,C12N5/00
CC Human gene and gene expression product v
FH Key Location/Qualifiers
FT source 1..872
FT /organism='Homo sapiens (human)'.
FEATURES
source 1..872
/organism='Homo sapiens'
/mol_type='genomic DNA'

/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 1..12e-14 Length: 872
Score: 1076.00 Matches: 216
Percent Similarity: 82.62% Conservative: 36
Best Local Similarity: 70.82% Mismatches: 34
Query Match: 8.14% Indels: 20
DB: 6 Gaps: 3

US-09-522-753-5 (1-2517) x BD221548 (1-872)

Qy 241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyAsnGlnPro 260
Db 10 GCTGTCGAAATTTTGAAGGTCTTGGCCCAAAAGTTGCACTGCCACTGTATACCAAGCCA 69
Qy 261 SerAspThrArgGlnTyHisGluAsnIleGluAsnIleGluAsnGlnAlaMetArgLysLeu 280
Db 70 TCAGATACCAAGGTGTACCATGAGAACATCAAGACCAACCAAGGTGTATGAGGAAAAAATC 129
Qy 281 IleLeuTyrrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
Db 130 ATTTTATTTTAAAGAGAAATCATGCAAGAAAAACAAAGGGAACAAAAAATCTGCAG 189
Qy 301 ArgTyrrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
Db 190 CGTTATGATCAGCTCTGAGGAGCATGGAGAAAAAGTGGACAGAAATAGAAAAATATCT 249
Qy 321 ArgArgAlaLysGluSerLysValArgGluTyrr-Tyr-GluLysGlnPheProGluIl 340
Db 250 CGGAGGAAAGCTAAAGAAACAAAAACWAGGGAATACTATTTRAAAAGCAGATTTCAGAAAT 309
Qy 340 eArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValClyGlnArgGlySerGl 360
Db 310 TCGAAACAAAGAGAACAGCAGAGAAAGATTTCAG---CGAGTTGGGAGAGGGAGCTGG 366
Qy 360 YLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSe 380
Db 367 TCITTCAGCCACCATTGCTAGGAGTGAGCATGAGATTCTCGAAATATTATGATGGCTCTC 426
Qy 380 rGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTy 400
Db 427 TGAGCAGGAGATAATAGAAACAAATTCGCGCAGCTCTC-GTGATTCCACCTATGATGTT 485
Qy 400 rAspAlaAspGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPome 420
Db 486 TGATGAGAACAAAGACGATCAAGTYCATTAACATGAATGGGCTTATGGAGGACCTAT 545
Qy 420 tLysValTyrrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPh 440
Db 546 GAAAGTGATAAAGATAGGCAGTTTATGAATGTTTGGACTGACCATCAAAAGGAGATCTT 605
Qy 440 eArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluAr 460
Db 606 TAAGGACAAGTTTATTCAGCATCCAAAAAATTTGGACTAATTCATCATCATCTTGGAGAG 665
Qy 460 gLysThrValAlaGluCysValLeuTyrrTyrrTyrrTyrrTyrrTyrrTyrrTyrrTyrr 480
Db 666 GAAGAGTGTCCTGATTGTTTGTATTACTATTTAAACCAAGAAAAAATGAGAAATTATAA 725
Qy 480 sSerLeuValArgArgSerTyrr---ArgArgArgGlyLysSerGlnGlnGlnGlnGln 499
Db 726 AGCCTCGTCAGAGGAATTTATGGAAACCGCAGAGGAGAAACCAAGCA-----774
Qy 499 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 519
Db 775 -----ATTGCTCGACCTCGCAAGGAAGA 797
Qy 519 uLysAspGluLysGluLysGluAlaGluLysGluGluLysGluGluLysProGluValGl 539
Db 798 AAAAGTAGAAGAAAAAGAGAGGATTAAGCAGAGAAAAACAAAAAAGAGAGAGAAAA 857

```
QY      539 uAsnAspLysGlu 543
      : |||::|||
Db      858 GAAAGATGAAG 870

RESULT 38
AX677866
LOCUS      AX677866          555 bp      DNA      linear      PAT 27-MAR-2003
DEFINITION Sequence 644 from Patent WO02086122.
ACCESSION AX677866
VERSION   AX677866.1 GI:29335271
KEYWORDS
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS   Legrain, P. and Daviet, L.
TITLE     Protein-protein interactions in adipocytes
JOURNAL   Patent: WO 02086122-A 644 31-OCT-2002;
          Hybrigenics (FR)
FEATURES
   source          1..555
                   /organism="Homo sapiens"
                   /mol_type="unassigned DNA"
                   /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:      1,72e-12      Length:      555
Score:          960.00      Matches:      184
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    7.26%      Indels:      0
Db:             6      Gaps:      0

US-09-522-753-5 (1-2517) x AX677866 (1-555)

QY      138 AspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHisThrAsp 157
      : |||::|||
Db      2 GACCGTAGCCTGACGGGAAAGCTGGAACCGGTGTCTCCCCAGCCCGCCGACACTGAC 61

QY      158 ProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuGluLeuGlnAsnMetAsp 177
      : |||::|||
Db      62 CTTGAGCTGGAGCTGGTCCCGCCAGCGGTGTCCAGAGGAGAGCTGATCCAGAACATGGAC 121

QY      178 ArgValAspArgGluLeuThrMetValGluGlnGlnLeuSerLysLeuLysLysGln 197
      : |||::|||
Db      122 CGCGTGGACCGAGAGATCACCATGCTAGAGCAGCAGATCTCTAAGCTGAAGAAAGCAG 181

QY      198 GlnGlnLeuGluGluAlaAlaLysProProGluLysProValSerProPro 217
      : |||::|||
Db      182 CAACAGCTGGAGGAGGCTGCCAAGCGCCCGAGCTGAGAGCCGTGTACCCCGC 241

QY      218 ProIleGluSerLysHisArgSerLeuValGlnIleLeuTyrAspGluAsnArgLysLys 237
      : |||::|||
Db      242 CCATCGAGTCGACGACCGCAGCTGGTGAGATCATCTACGACGAGAACCGGAAGAAG 301

QY      238 AlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyr 257
      : |||::|||
Db      302 GCTGAAGCTGCATCGGATCTGGAAGCCCTGGGGCCCGAGGTGGAGCTGCCGCTGAC 361

QY      258 AsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArg 277
      : |||::|||
Db      362 AACACGCCCTCCGACACCCCGGAGTATCATGAGACATCAAAATTAACACGCGCATGGG 421

QY      278 LysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGlnLys 297
      : |||::|||
Db      422 AGAAGCTAATCTTGTTACTTCAAGAGAGGAATCACGCTCGGAAACAATGGAAGCAGAAG 481

QY      298 PheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGlu 317
      : |||::|||
Db      482 TTCTGCAGCGCTATGACCAGCTCATGGAGGCGCTTGGAAAAAAGGTGGAGCGCATCGAA 541
```

```
QY      318 AsnAsnProArg 321
      : |||::|||
Db      542 AACACCCCGCGC 553

RESULT 39
HSU80750
LOCUS      HSU80750          560 bp      mRNA      linear      PRI 18-DEC-1997
DEFINITION Homo sapiens CTG26 mRNA, partial cds.
ACCESSION U80750
VERSION   U80750.1 GI:2565072
KEYWORDS
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS   Margolis, R.L., Abraham, M.R., Gatchell, S.B., Li, S.H., Kidwai, A.S.,
           Breschel, T.S., Stine, O.C., Callahan, C., McInnis, M.G. and Ross, C.A.
TITLE     cDNAs with long CAG trinucleotide repeats from human brain
JOURNAL   Hum. Genet. 100 (1), 114-122 (1997)
MEDLINE   97369492
PUBMED    9225980
AUTHORS   Margolis, R.L., Abraham, M.R., Gatchell, S.B., Li, S.H., Kidwai, A.S.,
           Breschel, T.S., Stine, O.C., Callahan, C., McInnis, M.G. and Ross, C.A.
TITLE     Direct Submission
JOURNAL   Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of
           Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
FEATURES
   source          1..560
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /tissue_type="cerebral cortex"
                   <1...>560
                   /gene="CTG26"
                   /note="CTG26"
                   /note="similar to mouse nuclear receptor co-repressor
                   encoded by GenBank Accession Number U35312; polyglutamine
                   rich"
                   /codon_start=3
                   /product="CTG26"
                   /protein_id="AAB91446.1"
                   /db_xref="GI:2565073"
                   /translation="VMMWSEQKETFRFKMHPKPNFLIASFLERKTVAECVLYYY
                   LKKENYKSLVRSYRRRQKQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
                   EKEEKPEVNDKEDLLKEKTDTSGEDNDEKAEVASKRKTANSQGRKGRKTRMSA
                   NEANSEEAITPQOSAEASMELENEP"
                   201..251
                   /genes="CTG26"
                   /notes="region encoding polyglutamine"
                   213..251
                   /rpt_type=tandem
                   /rpt_unit="CAG"

misc_feature
repeat_region

ORIGIN
Alignment Scores:
Pred. No.:      6.4e-12      Length:      560
Score:          932.00      Matches:      183
Percent Similarity: 99.46%      Conservative: 0
Best Local Similarity: 99.46%      Mismatches: 1
Query Match:    7.05%      Indels:      0
Db:             9      Gaps:      0

US-09-522-753-5 (1-2517) x HSU80750 (1-560)

QY      428 ValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHis 447
      : |||::|||
Db      3 GTCATGAACATGTGAGTGGAGGAGGAGACCTTCGCGGAGAGATTTCATGCACAT 62

QY      448 ProlysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysVal 467
      : |||::|||
```

[illegible]

```

RESULT 40
HSU00761/c
LOCUS
DEFINITION Homo sapiens CTG26 alternate open reading frame mRNA, complete cds.
ACCESSION U00761
VERSION U00761.1 GI:2565090
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
TITLE cDNAs with long CAG trinucleotide repeats from human brain
JOURNAL Hum. Genet. 100 (1), 114-122 (1997)
MEDLINE 97369492
PUBMED 9225980
REFERENCE
AUTHORS 2 (bases 1 to 560)
Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of
Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
FEATURES
source
1..560
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="brain"
/dev_stage="fetus"
<1..560
/genes="CTG26"
<1..375
/genes="CTG26"
/note="cysteine rich"
/codon_start=1
/product="CTG26 alternate open reading frame"
/protein_id="AAB91452.1"
gene
CDS

```

```

/db_xref="GI:2565091"
/translation="SFSSMEASSALCWGMVMASSLLASLATERVMRPLRLPWLAVLR
LEATAFSSLSPEVSFSLRRSSLSFTSGSSFSASFSSFSFSSWLLRGMGRC
CCCCCCCCCCCCCWLLPRR"
repeat_region      303..341
                    /rpt_type=tandem
                    /rpt_unit="CTG"

ORIGIN
Alignment Scores:
Pred. No.:          6,4e-12          length:          560
Score:              932.00           Matches:        183
Percent Similarity: 99.46%           Conservative:    0
Best Local Similarity: 99.46%         Mismatches:      1
Query Match:        7.05%            Indels:          0
DB:                  9                Gaps:            0

US-09-522-753-5 (1-2517) x HSU080761 (1-560)

Qy      427  GlnValMetAsnMetTrpSerGluGlnGluLySerPheArgGluLySerPheMetGln 446
Db      554  CAGTTCATGAACATGTGGAGTGAGCAGGAGAAGAGACCTTCCGGGAGAAGTTTCATGCAG 495
Qy      447  HisProLySerPheGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 466
Db      494  CATCCCAAGAACTTTGGCCCTGATCGCATCATCTCTCGAGAGGAAGACAGTGGCTGAGTGC 435
Qy      467  ValLeuTyTrTyTrLeuThrTyLysLeuAsnGluAsnTyLysSerLeuValArgArgSer 486
Db      434  GTCTCTATTACTACCTGTCGTAAAGAAGATGAGAACTATTAAGAGCCTGGTGAGACGGAGC 375
Qy      487  TyrArgArgArgGlyLySerSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506
Db      374  TATCGGCCCGCGGCAAGACGACGACGACCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 315
Qy      507  GlnGlnGlnGlnProMetProArgSerSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 526
Db      314  CAGCAGCAGCAGCCCATGCCCGCAGCAGCAGGAGGAGAACCGAGGTGGAGAACCAAGGAGGAG 255
Qy      527  LysGluAlaGluLySerGluGluLySerProGluValGluAsnAspLySerGluAspLeuLeu 546
Db      254  AAGGAGCGCGAGAAAGAGGAGGAGGAGAACCGGAGGTGGAGAACCAAGGAGGAGGAGGAGGAG 195
Qy      547  LysGluLySerThrAspAspThrSerGlyGluAspAsnAspGluLySerGluAlaValAlaSer 566
Db      194  AAGGAGNAGACACAGCACACCTCAGGGGAGGAGCAACGACGAGGAGGAGGAGGAGGAGGAGGAG 135
Qy      567  LysGlyArgLySerThrAlaAsnSerGlnGlyArgGlyArgGlyArgGlyArgGlyArgGlyMet 586
Db      134  AAAGGCGCAAACTGCAACAGCCAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 75
Qy      587  AlaAsnGluAlaAsnSerGluGluAlaLeuThrProGlnGlnSerAlaGluLeuAlaSer 606
Db      74   GCTAATGAGGCCAACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 15
Qy      607  MetGluLeuAsn 610
Db      14   ATGAGCTGAAT 3

RESULT 41
AX753058/c
LOCUS      AX753058              718 bp      DNA      linear      PAT 23-JUN-2003
DEFINITION Sequence 171 from Patent EP1310567.
ACCESSION  AX753058
VERSION     AX753058.1  GI:32165818
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Stuhlmueller,B. and Haeupl,T.
  TITLE     Nucleic acid array

```



```

Db 235 TTTGGACTAATTGCATCATCTACTTGGAGAGAGAGTCTTCTGATTGTTGTTGTTATAC 176
Qy 471 TylLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArg 489
Db 175 TATTTAACCAAGAAATGAGAAATATAAAGCCCTCGTCAGAGGAATTTATGGGAACGC 116
Qy 490 ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Db 115 AGAGCGAGAAACACGCAA----- 98
Qy 510 GlnProMetProArgSerGlnGluLysAspGluLysGluLysGluLysGluLysGluLa 529
Db 97 -----ATTGCTGCACCTCGCAAGAAAGAAAGTAAAGAAAGAGGATTAAGCA 44
Qy 530 GluLysGluGluGluLysProGluValGluAsnAspLysGlu 543
Db 43 GAAAAACACAGAAAAAAGAGAGAAAGAAAGAGATGAAGAG 2

RESULT 43
BD101314/c
LOCUS BD101314 710 bp DNA linear PAT 27-AUG-2002
DEFINITION Novel genes cloned in humanneuroblastoma and fragments thereof.
ACCESSION BD101314
VERSION BD101314.1 GI:22646888
KEYWORDS WO 0166719-A/3614.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Nakagawara,A.
AUTHORS Nakagawara,A.
TITLE Novel genes cloned in humanneuroblastoma and fragments thereof
JOURNAL CHIBA PREF.HISAMITSU PHARMACEUTICAL CO INC,AKIRA NAKAGAWARA
COMMENT OS Homo sapiens (human)
PN WO 0166719-A/3614
PD 13-SEP-2001
PF 02-MAR-2001 WO 2001JP001629
PR 07-MAR-2000 JP 00P 159195
PI AKIRA NAKAGAWARA
PC CL2N15/11,CL2Q1/68,G01N33/53,G01N33/566
CC Novel genes cloned in humanneuroblastoma and fragments thereof
FH Key Location/Qualifiers
FT source 1..710
FT /organism="Homo sapiens (human)".

FEATURES
source
1..710
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 4.46e-10 Length: 710
Score: 845.00 Matches: 174
Percent Similarity: 82.28% Conservative: 35
Best Local Similarity: 68.50% Mismatches: 27
Query Match: 6.39% Indels: 20
DB: 6 Gaps: 3

US-09-522-753-5 (1-2517) x BD101314 (1-710)

Qy 291 ArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluLaLeuGlu 310
Db 710 AGAAACAAAGGNCNCAAAATCTGCAGCGTTATGTCAGGCTCATGAAGCATG-GAA 652
Qy 311 LysLysValGluArgLysGluAsnAsnProArgArgArgAlaLysGluSerLysValArg 330
Db 651 AAAAAAGTGCAGATAGAAAATAT-CTTCGNAGGAAGCTAAAGAAAGCAACCAAGG 593
Qy 331 GluTyrTrpGluLysGlnPheProGluLysGlnArgGluLeuGlnGluArgMet 350
Db 592 GAATACTATGAAAGCGAGTTTCCAGAAATTCGAACCAAGAGAGACCAAGCAAGAGATTT 533

```

```

Qy 351 GlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHis 370
Db 532 CAG---CGAGTTGGCAGAGCGGAGCTGTTCTTCAGCCACCATTTGCTAGGAGTGAGCAT 476
Qy 371 GluValSerGluLeuLeuLeuAspGlyLeuSerGluGlnGlnGlnGlnGlnGlnGlnGln 390
Db 475 GRATTTCGAAATATTGATGGGCTCTCTGAGCAGGAGAAATATATGAGAACAAATCGG 416
Qy 391 GlnLeuAlaValIleProMetLeuTyrAspAlaAspGlnGlnArgGlyLeuPheIle 410
Db 415 CAGCTCTCTGATTCACCTATGATGTTGATGTCAGAACAAAGACGAGTCAAGTTCAAT 356
Qy 411 AsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsn 430
Db 355 AACATGAATGGCTTATGAGGAGCCCTATGAAAGTGATATAAGATAGGAGTATATGAAT 296
Qy 431 MetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsn 450
Db 295 GTTTGGACTACCATGAAAGAGAGATCTTTAAGACAAAGTTTATCCAGCATCCAAAAAC 236
Qy 451 PheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyr 470
Db 235 TTTGGACTAATTCATCATCTACTTGGAGAGAGAGTCTTCTGATGTTGTTGTTATAC 176
Qy 471 TyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArg 489
Db 175 TATTTAACCAAGAAATGAGAAATATAAAGCCCTCGTCAGAGGAATTTATGGGAACGC 116
Qy 490 ArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
Db 115 AGAGCGAGAAACACGCAA----- 98
Qy 510 GlnProMetProArgSerGlnGluLysAspGluLysGluLysGluLysGluLysGluLa 529
Db 97 -----ATTGCTGCACCTCGCAAGAAAGAAAGTAAAGAAAGAGGATTAAGCA 44
Qy 530 GluLysGluGluGluLysProGluValGluAsnAspLysGlu 543
Db 43 GAAAAACACAGAAAAAAGAGAGAAAGAAAGAGATGAAGAG 2

RESULT 44
AX677865 LOCUS AX677865 673 bp DNA linear PAT 27-MAR-2003
DEFINITION Sequence 643 from Patent WO2086122.
ACCESSION AX677865
VERSION AX677865.1 GI:29335270
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Legrain,P. and Daviet,L.
TITLE Protein-protein interactions in adipocytes
JOURNAL Patent: WO 02086122-A 643 31-OCT-2002;
Hybridomics (PR)
FEATURES
source
1..673
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 1.14e-08 Length: 673
Score: 774.50 Matches: 159
Percent Similarity: 80.09% Conservative: 14
Best Local Similarity: 73.61% Mismatches: 38
Query Match: 5.86% Indels: 5
DB: 6 Gaps: 3

US-09-522-753-5 (1-2517) x AX677865 (1-673)

```



```

repeat_region 1479. .1611 /rpt_family="L1"
repeat_region 1613. .1703 /rpt_family="Alu"
repeat_region 1725. .2102 /rpt_family="MaLR"
repeat_region 2103. .2233 /rpt_family="ERV1"
repeat_region 3447. .3809 /rpt_family="MER1_type"
repeat_region 3833. .3853 /rpt_family="AT_rich"
repeat_region 3877. .4201 /rpt_family="Alu"
repeat_region 4960. .5140 /rpt_family="MIR"
repeat_region 5225. .5308 /rpt_family="MIR"
repeat_region 5678. .6101 /rpt_family="ERV1"
repeat_region 7617. .7656 /rpt_family="AT_rich"
repeat_region 7904. .8270 /rpt_family="MaLR"
repeat_region 9980. .10090 /rpt_family="MIR"
repeat_region 10227. .10952 /rpt_family="L1"
repeat_region 10953. .11210 /rpt_family="L1"
repeat_region 11211. .11477 /rpt_family="L1"
repeat_region 11519. .11877 /rpt_family="MaLR"
repeat_region 11878. .12504 /rpt_family="L1"
repeat_region 12519. .12689 /rpt_family="Alu"
repeat_region 12702. .12737 /rpt_family="L1"
repeat_region 12738. .13049 /rpt_family="Alu"
repeat_region 13050. .13173 /rpt_family="L1"
repeat_region 13176. .13299 /rpt_family="MIR"
repeat_region 13327. .13486 /rpt_family="MIR"
repeat_region 16006. .16029 /rpt_family="MIR"
repeat_region 16449. .16693 /rpt_family="(TTA)n"
repeat_region 17836. .17950 /rpt_family="MIR"
repeat_region 18145. .18396 /rpt_family="Alu"
repeat_region 21611. .21715 /rpt_family="L1"
repeat_region 21716. .21751 /rpt_family="Alu"
repeat_region 22033. .22106 /rpt_family="AT_rich"
repeat_region 23436. .23459 /rpt_family="L2"
repeat_region 25281. .25333 /rpt_family="(T)n"
repeat_region 25643. .25665 /rpt_family="AT_rich"
repeat_region 25699. .25853 /rpt_family="AT_rich"
repeat_region 25876. .26149 /rpt_family="L1"
repeat_region 26156. .26177 /rpt_family="Alu"

```

```

repeat_region /rpt_family="(TAA)n"
27215. .27518 /rpt_family="Alu"
27674. .27981 /rpt_family="Alu"
27983. .28156 /rpt_family="ERV1"
28158. .28452 /rpt_family="Alu"
28453. .28760 /rpt_family="Alu"
28737. .29016 /rpt_family="ERV1"
29017. .29400 /rpt_family="ERV1"
29411. .29522 /rpt_family="L1"
29524. .29783 /rpt_family="Alu"
30116. .30417 /rpt_family="Alu"
31701. .31897 /rpt_family="MIR"
32032. .32330 /rpt_family="Alu"
34009. .34503 /rpt_family="L1"
35112. .35223 /rpt_family="L1"

```

Alignment Scores:

```

Pred. No.: 1.36e-06 Length: 135351
Score: 763.50 Matches: 606
Percent Similarity: 32.45% Conservative: 339
Best Local Similarity: 20.81% Mismatches: 1284
Query Match: 5.78% Indels: 693
DB: 9 Gaps: 106

```

US-09-522-753-5 (1-2517) x AC068279 (1-135351)

```

QY 2 SerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProPro 21
Db 101835 GCTGGCCAAACC-----AGCCAAAGCCAGTCAGCAAGCCAGCCAGCCAGCCACCC 101885
QY 22 HisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspVal-----Gly 38
Db 101886 AGCCAAAGCCAGCAAGCCAGCCAAACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 101945
QY 39 LeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySer--- 57
Db 101946 AGCCAAAGAGACCCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 102005
QY 58 -----IleIleGlnPro-----GlnArgArgArgProSerLeuLeu 69
Db 102006 CCAGCCAAAGCCGGCCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 102065
QY 70 SerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHis 89
Db 102066 CCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 102125
QY 90 SerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArg 109
Db 102126 GCCAGCCAGCCAGACACA-GGCCAGCCAAAGCAAGCCATGGAAGCCAGCCAGCCAGCCAGCC 102181
QY 110 LeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPro 129
Db 102182 ---GCCAAGCTAGCCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 102238
QY 130 AlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSer 149
Db 102239 -----AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 102292
QY 150 Pro-ProSerProProHisThrAspProGluLeuLeuValProProArgLeuSerLys 169

```


[illegible][illegible]

[illegible]

Query Match:	5.68%	Indels:	63
DB:	9	Gaps:	8

US-09-522-753-5 (1-2517) x AK057740 (1-1793)

Qy	524	GlulysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGlu	543
Db	1	GAGAAGAAAGAGCGCGAAGAGGAGGAAGCGGAGTGGAGAACGACAAGAA	60
Qy	544	AspLeuLeuLysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAla	563
Db	61	GACTCTCTCAAGAGAACACAGACGACACCTCAGGGGAGCAACGACGAGAGAGCGCT	120
Qy	564	ValAlaSerLysGlyArgLysThrsAlaAsnSerGlnGlyArgArgLysGlyArgLleThr	583
Db	121	GTGGCCCTCCAAAGGCCCAAATGCACACAGCCAGGGAAGACCAAGGCGCATCAC	180
Qy	584	ArgSerMetAlaAsnGluAlaAsnSerGluGluAlalleThrProGlnGlnSerAlaGlu	603
Db	181	CGCTCAATGGCTAATGAGGCCAAAGAGGAGGCGCATCACCCCACGAGAGCGCCGAG	240
Qy	604	LeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAla	623
Db	241	CTGGCCTCCATGGAGCTGAATGAGATTCTCGTTGGA CAGAAAGAAAATGGAACAGCC	300
Qy	624	LysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlalleAlaAacMetValGlySer	643
Db	301	AAGAAAGGTCTCCTGGAAACACGGCGCCAATGGTCGCCCATCGCCCCGATGGTGCGCTCC	360
Qy	644	LysThrValSerGlnCysLysAsnPheTyxPheAsnTyrlsLysArgGlnAsnLeuAsp	663
Db	361	AAGACTGTGTGCGAGTGAAGAACTTCTACTTCACTACAGNAGAGGCGACAACTTAAGG	420
Qy	664	GluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLys	683
Db	421	AGGITCTCGGACCCCGAGCTGTGGCCCTGGGGCGCGGGCGTCTGTTGGGTGCTGTGTCG	480
Qy	684	LysLysAlaProAlaAlaSerGluGluAlaAlaPhePro-----ProVal	699
Db	481	TTGAGGGTGTGTGGCGTGTGTGGC---GTCTTAGCTGTGCCCTCGCCCTGGTGTGTCGCCACGG	537
Qy	700	ValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGlu-MetValGlu	719
Db	538	ACGAGTGACCCCTCTGAGCGCTCGGAGCCCGGAATAGCAGCTGGGAGAGGGCGGA	597
Qy	719	uGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyPr	739
Db	598	GGACTGAGCTCGCTGCCCGCTCGCGGCGG-----	628
Qy	739	oAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAl	759
Db	629	-----GC 630	
Qy	759	alysAspThrGlyGlnAsnGlyProlysProAlaThrLeuGlyAlaAspGlyPro---	778
Db	631	TAGAGATAAG-----GGCGTGGCCCTTGTGTGCGCTCCCGCCCTCGGTCCCT	681
Qy	779	-----ProProGlyPro-----ProThrProProArgArgThrSerArgAlaPr	793
Db	682	CACCTCGGGCCACCTCGGGCGGTGCTCTGCCCCGCGACAGTAGCCCGCTCGGA	741
Qy	793	oIleGluProThrProAlaSerGluAlaThrGlyAlaProThrPro-----	808
Db	742	GCGAGGCCCGCTCACCTGGTGCTGTACTACTCCCCGGGCGGTGGGCGAAGACA	801
Qy	809	-----ProProAlaProProSerProSerAlaProPr	819
Db	802	CCCGCAGGAACCTCGCAGAGGAGAAATTACAGCGGTCCCGAGGGTAGGAAAAAGACCCCG	861
Qy	819	oProValValProlysGluLysGluGlu-----ThrAlaAlaAl	839
Db	862	GCCACCGTGAATCTGAAACACCCGACCACCTCGGCATCCCATGTTTTCACGAGTCAGAC	921

```
Qy 834 aproProValGluGluGlyGluGlu 842
Db 922 CCCAGGCGCAGGCGAGGCGAGGAGCAG 946

RESULT 48
AX396270 LOCUS 520 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 485 from Patent WO0212328.
ACCESSION AX396270
VERSION AX396270.1 GI:21067017
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS King G.E., Meagher M.J., Xu, J. and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon
JOURNAL Patent: WO 0212328-A 485 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
source 1..520
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 3.41e-08 Length: 520
Score: 746.50 Matches: 141
Percent Similarity: 91.95% Conservative: 19
Best Local Similarity: 81.03% Mismatches: 13
Query Match: 5.65% Indels: 1
DB: 6 Gaps: 1

US-09-522-753-5 (1-2517) x AX396270 (1-520)

Qy 285 LysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGln 304
Db 1 AAGAAGAGAAATCATGCGAGAAACAAAGGGAACAAAATCTGCCAGCGTTATGATCAG 60

Qy 305 LeuMetGluAlaLeuGluLysLysValGluArgGlnGlnPheProGluLysGlnArg 324
Db 61 CTATGAGGCGATGGGAGAAAAAGTGCACAGATAAATAATCTCTCGAGGAGAAAGCT 120

Qy 325 LysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluLysGlnArg 344
Db 121 AAGAAGAGCAAAACAGGGAATATATGAAAGCAGTTTCCAGAAATTCGAAACCAAGA 180

Qy 345 GluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSer 364
Db 181 GAACAGCAAGAGATTTTCAG--CGAGTTGGCGAGAGGAGGCTGCTTTTCAGCCACC 237

Qy 365 AlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsn 384
Db 238 ATTGCTAGGAGTGAGCATGAGATTCTGAAATTATTGATGGCTCTCTGAGCAGGAGAA 297

Qy 385 LeuGluLysGlnMetArgGlnLeuAlaValIleProPheMetLeuTyrAspAlaAspGln 404
Db 298 AATGAGAAACAAATGCGCGAGCTCTCTGTGATTCCACCTATGATGTTGATGAGAACAA 357

Qy 405 GlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys 424
Db 358 AGACGAGTCAAGTTTCATTAATCATGATGATGGCTTATGAGGACCCCTATGAAGGTATPAA 417

Qy 425 AspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPhe 444
Db 418 GATAGGCAGTTTATGATGTTTGGACTGACCATGAAAGGAGATCTTTTAAGGACAAAGTTT 477

Qy 445 MetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeu 458
Db 478 ATCCAGCATCCAAAACTTTGGACTAATTTGCATCATCTTG 519
```

RESULT 49

```
AC020019/c LOCUS 40871 bp DNA linear HTG 03-JAN-2000
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
```

```
AC020019 ACCESSION
AC020019.1 GI:6664878
HTG; HTGS_PHASE2.
```

```
KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
```

```
REFERENCE 1 (bases 1 to 40871)
AUTHORS Adams, M. and Venter, J. C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
```

```
COMMENT This sequence was identified as CDM:10211762 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
```

```
FEATURES
source 1..40871
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
```

ORIGIN

```
Alignment Scores:
Pred. No.: 1.34e-06 Length: 40871
Score: 743.00 Matches: 237
Percent Similarity: 41.08% Conservative: 106
Best Local Similarity: 28.38% Mismatches: 254
Query Match: 5.62% Indels: 238
DB: 2 Gaps: 23
```

```
US-09-522-753-5 (1-2517) x AC020019 (1-40871)
```

```
Qy 144 LysLeuGluProValSerProProSerProProHisThrAspProGluLeuGluLeuVal 163
Db 11990 CAGGTGGAGGCCATTTTCGCCGACATGCCCC-----AGCGATAGTTTCGATTGAGGCGC 11937
```

```
Qy 164 ProProArgLeuSerLysGluGluLeuGlnAsnMetAspArgValAspArgGluLeu 183
Db 11936 GGAAGGACCGCCGCAAGAGGATCTCTCATCAATCCAAAGGTTGACATGAGATC 11877
```

```
Qy 184 ThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluGlu 203
Db 11876 AATCCGCTGAGACGACAAATGGAAACGTTGCGCAAAAGGAGAAATCCCTCATGAGGAG 11817
```

```
Qy 204 Ala-----AlaLysProGluProGluProGluProValSer----- 215
Db 11816 GCGCGCTGCGCCAAAGAGGAGGCGCGCCAAAGAGTTGAACCAATCAACAATGATCAG 11757
```

```
Qy 216 ProProIleGluSerLysHisArgSer-----LeuValGlnIleIleTyrAspGlu 233
Db 11756 GAACCATGATGTGAACATCTGTCGGCGAGCCAAATGCTACGGGAGAGATCTATGCGGCC 11697
```

```
Qy 234 AsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGlu 253
Db 11696 AATCGGAAGAGCGCCCAAGCGCAACATTCATCTGTCAGATGATCAGCGCGCGATCG 11637
```

```
Qy 254 -----LeuProLeuTyrAsnGlnProSerAspThr 263
Db 11636 TCTCCGGGTTCCGTTGTCGGCGACACCATGTTGCCATTGTATTAACCAAGCCTCATGTC 11577
```

```
Qy 264 ArgGlnTyrHisGluAsnIleLysIleAsnGlnAla---MetArgLysLysLeuIleLeu 282
Db 11576 GAGGCATCTTGCCCATGTGTAATACCGCAGCATCAGAGTCAAAATCGTCGCGCTACTTTTG 11517
```

QY	283	TyrPheLysArgArgAsnHisAlaAaGlyGlnTrpLysGlnLysPheCysGlnArgTyr	302	Db	10653	-----	10653
Db	11516	CAATACGAAATTTGAAGCCGACCGTGGCGCACAATCAGGAGCTGTGGAGAGTAC	11457	QY	646	ValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle	665
QY	303	AspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnProArgArg	322	Db	10653	-----	10653
Db	11456	ACCAAGATCAGCGGATTTGGCAGCAGCATGCGAGGTATGGAGCCAGTGCAGAGG	11397	QY	666	LeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaAaArgLysLysLys	685
QY	323	ArgAlaLysGluSerLysValArgGluTyrGluLysGlnPheProGluIleArgLys	342	Db	10652	-----	10632
Db	11396	AAAGCAGCGGAGGCAAAATCGGAATCTTCCGAAAGGCTCTTACGAGGATTAAGGAA	11337	QY	686	AlaProAlaAlaAsnSerGluGluAlaAlaPheProValValGluAspGluMet	705
QY	343	GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer	362	Db	10631	GCGGTGCGAGCGCGCGCGCAGCAGCAACTACCGCCACACCGCGCAACCAACCA	10572
Db	11336	CAGCGCAGGATAAGGAACGTTTC--AATCGCGTGGTTCGCGG-----	11295	QY	706	GluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHis	725
QY	363	MetSerAlaAlaArgSerGluHisGluValSerGluIleLeuAspGlyLeuSerGluGln	382	Db	10571	TCATCATCAACATCATCATCATCATCA	10542
Db	11294	-----ATTAAATCCGAAGCGGATCTCGAGGAGATCATGATGCGCTGCAGGAGCAG	11244	QY	726	AlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSer	745
QY	383	GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAla	402	Db	10541	GCATCATCAGCATCAACAGCATCATCGCGCAAGTCCAGCGACATTG	10491
Db	11243	CGGTGAGGATAAAAGATCGCATCTATCGCGTTATACCGCTCTGATGCAAGTGC	11184	QY	746	SerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn	765
QY	403	AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVal	422	Db	10490	-----GCAGGAATAGCCGACAGATAAACCGATGCCGCAAA--ACGCTAGCGCC	10443
Db	11183	CGCCAGAGGCGATCGCGCTATCATATGAATGGACTTATCGAGGATATGTTGGCAGTT	11124	QY	766	GlyProLysProProAlaThrLeuGlyAlaAspGlyProProProGlyProProThrPro	785
QY	423	TyrLysAspArgGlnValMetAsnMetTyrSerGluGlnGluLysGluThrPheArgGlu	442	Db	10442	ACGGTAAGATGTCAGCAACAGCTGT-----GGACCA-----	10410
Db	11123	CATCAACAGCGTAAGGCCCTCAACATGTGGACCGCGCGAGAGGAGACCTTTAAGGAG	11064	QY	786	ProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAla	805
QY	443	LysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThr	462	Db	10409	-----ACGCCAACGGGAAT	10395
Db	11063	AAGTATTTCAGCATCCGAAATTTGGAGCCATTCGCGCTAGTTTGGATCGCAATCG	11004	QY	806	ProThrProProAlaProProSerProSerAlaProProProValValProLysGlu	825
QY	463	ValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu	482	Db	10394	CGACAGCAGCAACTACCTGCAACAGCAACAGCACCA	10353
Db	11003	CCACAGGATCGTTCGTACTACTACTCTCAGCAAGAGACCGAGAGATTACAAGCAGCTG	10944	QY	826	GluLysGluGluGluThrAlaAlaProProValGluGluGluGluGluGlnLysPro	845
QY	483	ValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln	502	Db	10352	-----GAGATCTCAGCGGC-----GGCAGCGCAAGAGCAAA	10320
Db	10943	CTTCGAAATCCCGCAACGACGCGAGTTCGCGTAAATCCAGCAAGGCTCAGCGCGCA	10884	QY	846	ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGlu	865
QY	503	Gln-----GlnGlnGln	506	Db	10319	AATGCGAGGAGGAGCAGCTGCCAGCGGAGCAGCA-----	10281
Db	10883	CAGCGCAGTGATTCATTCGATGATGACCTGCGGTTATGCGCGATTCAGCGCGGAA	10824	QY	866	CysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAla-----Glu	881
QY	507	GlnGlnGlnGlnProMetProArgSerSerGln---GluGluLysAspGluLysGluLys	525	Db	10280	-----ACGGTGGCCACTCCGGTACACCGCAACCGGAGCAAGTCCGCGCAGTGTGCGGAG	10224
Db	10823	CAGCAACAAAGTCCGCGGTGTTTCATCCCGGTGCTGAGCGTGCAGCGTGCAGAACT	10764	QY	882	AlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGly	901
QY	526	GluLysGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu	545	Db	10223	GCGACAACGCGCAACGCGCGCCAGCTGCGCGCAAGGGTGTTCGCAAAACCGGAA	10164
Db	10763	CGACCGCAGCGTGCAGCGTGTGCGAGAGAAAGCGCAGCTGATGTGTCGCAAA	10713	QY	902	ArgAla-----ThrThrAlaLysSerSerGlyAlaProGlnAspSerAsp	916
QY	546	LeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaValAla	565	Db	10163	ACAGCGAGTGCAGCGGAAACGGTGTCTAAGGAGAGTACTCTCGGCAGATGCAAC	10104
Db	10712	-----CGGCTGAAAGTGCAGCGGAAAGCGCAAGCGCGCC	10677	QY	917	SerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGly	931
QY	566	SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSer	585	Db	10103	GATCCGCTGCGCAAACTGTTCTGAAGGCTATCAATGCCGAGGCG	10059
Db	10676	ACTAAG-----	10671	RESULT 50			
QY	586	MetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAla	605	AC023741/c			
Db	10670	-----GCTGTCGAGCAACGCT-----	10653	LOCUS			
QY	606	SerMetGluLeuAsnGluSerSerArgTyrThrGluGluGluMetGluThrAlaLysLys	625	DEFINITION			
Db	10653	-----	10653	AC023741			
QY	626	GlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThr	645	AC023741.4			
				GI:21450410			
				HTG.			
				KEYWORDS			
				Drosophila melanogaster (fruit fly)			
				SOURCE			
				ORGANISM			
				Drosophila melanogaster			
				INSTRUMENT			
				160440 bp			
				DNA			
				linear			
				INV 18-JUN-2002			
				Drosophila melanogaster X BAC RP98-2309 (Roswell Park Cancer			
				Institute Drosophila BAC Library) complete sequence.			

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 160440)

REFERENCE AUTHORS

Muzny, D., Scherer, S., Adams, M.D., Holt, R.A., Evans, C.A., Gocayne, J.D., Tabot, P., Williamson, A., Homs, F.H., Dugan-Rocha, S., Sodergren, E.S., Hodgson, A.H., Chen, R.C., Ayele, M., Scott, G.S., Worley, K.W., Amaratunga, H.C., Brannon, R.C., Rogers, Y., An, H., Baldwin, D., Beeson, K.Y., Brown, M., Buhay, C., Busam, D.A., Center, A., Chen, G., Chen, Z., Clerc-Blankenhay, K., Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E., Draper, H., Emery-Cohen, A., Ferrera, S., Garg, N.D.S., Houck, J., Hostin, D., Howland, T.J., Hume, J., Ibegwam, C., Jalili, M., Kovar, C., Liu, W., Mattei, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B., Nelson, K.A., Ndassa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V., Scheeler, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M., Xiang, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O., Wheeler, D., Weinstock, G., Gibbs, R. and Venter, J.C.

Direct Submission

TITLE JOURNAL

Unpublished

REFERENCE AUTHORS

2 (bases 1 to 160440)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Bimge, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loujsegh, H., Lozado, R., Mapua, P., Martin, R., Martindale, A., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojuben, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

Direct Submission

TITLE JOURNAL

Unpublished

REFERENCE AUTHORS

3 (bases 1 to 160440)
Worley, K.C.
Direct Submission
Submitted (17-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 160440)

REFERENCE AUTHORS

Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Bimge, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loujsegh, H., Lozado, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojuben, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

Direct Submission

TITLE JOURNAL

Submitted (01-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 160440)

REFERENCE AUTHORS

Submitted (22-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

6 (bases 1 to 160440)

REFERENCE AUTHORS

Submitted (18-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jun 19, 2002 this sequence version replaced gi:18030103.

COMMENT

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not

identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.htm>.

FEATURES

source

Location/Qualifiers

```
1.160440
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="X"
/clone="RP98-2309"
1846..1857
/note="Baylor,12Cs/Celera,13Cs"
/function="polymorphic site"
60056..60059
/function="low quality"
100089
/note="C/T?"
/function="unSURE"
153760..156660
/note="Confirmed by restriction digest."
/function="unresolved tandem repeat"
155589..155590
/function="low quality"
```

ORIGIN

Alignment Scores:
 Pred. No.: 4,04e-06 Length: 160440
 Score: 743.00 Matches: 237
 Percent Similarity: 41.08% Conservative: 106
 Best Local Similarity: 28.38% Mismatches: 254
 Query Match: 5.62% Indels: 238
 DB: 3 Gaps: 23

US-09-522-753-5 (1-2517) x AC023741 (1-160440)

```
QY 144 LysLeuGluProValSerProSerProHisThrAspProGluLeuVal 163
DB 72241 CAGGTGGAGGCCATTTCGCCGACATTGCC-----ACGATAGTTGAGGAGGCG 72188
QY 164 ProProArgLeuSerLysGluLeuLeuLeuLeuLeuLeuLeuLeuLeu 183
DB 72187 GGAAGGACCGCCGACCAAGAGGATCTCTCATGCAATCCAAAGGTTGACAAATGAGATC 72128
QY 184 ThrMetValGluGlnGlnLeuSerLysLeuLysLysGlnGlnGlnGlnGlu 203
DB 72127 AAATCCGCTGAGACGCAATAGAAACGTTGCGCAAAAGAGAAATCCCTCATGAGGAG 72068
QY 204 Ala-----AlaLysProGluProGluLysProValSer----- 215
DB 72067 GCGCCCTGCGCAAGAGCAGAGCGCGCCCAAGAGTTGAACGATAACAAATGATCAG 72008
QY 216 ProProProLeuGluSerLysHisArgSer-----LeuValGlnLeuLeuLeuVal 233
DB 72007 GAACCGATGTTGAACTATCTGCGCGCAGCAATGCTAGCGGAGAGATCTATGCGGCC 71948
QY 234 AspArgLysLysAlaGluAlaAlaHisArgLeuLeuGluLeuGlyLeuGlyProGlnValGlu 253
DB 71947 AATCGAAGACGCGCCCAAGCGCAACATTCATGCTGAGATCGAGCGCGCGATGAATCG 71888
QY 254 -----LeuProLeuTyraSngLnnProSerAspThr 263
```

```
DB 71887 TCTCCGGTTCCTGTCGGCGCAGACCATGTTGTCATTGTATAACCGACCTCGATGTC 71828
QY 264 ArgGlnTyHisGluAsnLeuLysLeuAsnGlnAla---MetArgLysLysLeuLeuLeu 282
DB 71827 GAGGCATTCGCCATGCTTAATACGCCAGCATCAGATCAAAATCTGTCGGCGCTACTTTTG 71768
QY 283 TyrPhelysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyr 302
DB 71767 CACATAGAAATTTGAAGCCGCAACCTGGCGCCACAAATCAGGACTGGTGGAGAAGTAC 71708
QY 303 AspGlnLeuMetGluAlaLeuGluLysLysValGluArgLysLeuAsnProArgAG 322
DB 71707 ACCAAGATCAGGGCGATTGGCAGCAGCATGCGAGCGTATGGAGGCCAGTCCACAGAG 71648
QY 323 ArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluLeuLeuArgLys 342
DB 71647 AAGCAGCGAGGCAAAAATTCGCGAATCTTTCGAAAGGTCTTCACGGAGCTAAGGAAA 71588
QY 343 GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer 362
DB 71587 CAGCGCAGGATAGGAACGTTTC---AATCGCTCGGTTTCGCG----- 71546
QY 363 MetSerAlaAlaArgSerGluHisGluValSerGluLeuLeuLeuLeuLeuLeuLeu 382
DB 71545 -----ATTAAATCCGAAGCGGATCTCGAGAGATCATGGAATGCGCTTGATCAGCATGCC 71495
QY 383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValLeuProProMetLeuTyraSAla 402
DB 71494 GCGTTGGAGGATAAAGATGCGATCCTATCGGTTATACGCTCTGATCAGCATGCC 71435
QY 403 AspGlnGlnArgIleLysPheLeuAsnMetAsnGlyLeuMetAlaAspProMetLysVal 422
DB 71434 CCGCAGAGCGCATCGCCTATCATATAAATGAGATTCAGGAGTATGTCGGCGATT 71375
QY 423 TyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGlu 442
DB 71374 CATCAACAGCGTAAGGCCCTCAACATGTGGACCGCGCGAGAGAGACCTTTAAGGAG 71315
QY 443 LysPheMetGlnHisProLysAsnPheGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 462
DB 71314 AAGTATTCAGCATCCGAAAAAATTTGGAGCAATGTCGGCTAGTTTGATCGCAATTCG 71255
QY 463 ValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu 482
DB 71254 CCACAGGATTGCGTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 71195
QY 483 ValArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGln 502
DB 71194 CTTGCAAAATCCGCCAACCGGACGCGCAGTTCGCGTAATCCAGCCAAAGGCTCAGCGCA 71135
QY 503 Gln-----GlnGlnGln 506
DB 71134 CAGCGCGAGTGCATTATCGATTCGATGCACCTGGCGTTATACGCGATTGACGCGCGAA 71075
QY 507 GlnGlnGlnGlnProMetProArgSerSerGln---GluGluLysAspGluLysGluLys 525
DB 71074 CAGCAACAAAGGTCGCGCGCTCTCATCGCGCTGCTGAGCGTTCGCGTTCGCGAACGT 71015
QY 526 GluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLeu 545
DB 71014 GCGCGCGAGCGTGAGCGTGTGGCAGAGAAAGCGGACGCTGCTGCTGCCAAA----- 70964
QY 546 LeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaValAla 565
DB 70963 -----GCGGCTGAAGTGCCTCCGCCGCAAGGCAAGCGCGCC 70928
QY 566 SerLysGlyArgLysThrAlaAsnSerGlnGlyArgLysGlyArgLysThrArgSer 585
DB 70927 ACTAAG----- 70922
QY 586 MetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAla 605
```

Db	70921	---GCTGTGGAAGCAACGGCT---	70904
Qy	606	SerMetGluLeuAAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLys	625
Db	70904	-----	70904
Qy	626	GlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThr	645
Db	70904	-----	70904
Qy	646	ValSerGlnCysLysAsnPheTyPheAsnTyRlyLysArgGlnAsnLeuAspGluIle	665
Db	70904	-----	70904
Qy	666	LeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgGlyLysLysLys	685
Db	70903	-----GCGGTGAAAGGTGCGCAAA	70883
Qy	686	AlaProAlaAlaAspGluAlaAlaPheProProValValGluAspGluMet	705
Db	70882	GCGCTGCGAGCGCGCGCGCGAGCACTACCGCCACACCAACCAACCA	70823
Qy	706	GluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHis	725
Db	70822	TCATCATCAACATCATCATCATCA	70793
Qy	726	AlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSer	745
Db	70792	GCATCATCAGCATCAACAGCATCATCATCGAGCGAGTCCAGCGACATG	70742
Qy	746	SerAspThrGluSerIleProSerProHisThrGluAlaLysAspThrGlyGlnAsn	765
Db	70741	---GCGAATAGCGCGAGATAAAGCATGCGCGCAAA---ACGGTAGCGCC	70694
Qy	766	GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro	785
Db	70693	AGCGATAAGATGAGCAACAGCTGGT	70661
Qy	786	ProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAla	805
Db	70660	---ACGGCACCGGGAAGT	70646
Qy	806	ProThrProProAlaProProSerProSerAlaProProProValProLysGlu	825
Db	70645	CCGACAGCACTACCTGCAACAGCAACAGCAACCA	70604
Qy	826	GluLysGluGluThrAlaAlaProProValGluGluGluGluGluGluGluPro	845
Db	70603	---GAGATCTCAGCGGC---	70571
Qy	846	ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGlu	865
Db	70570	AATGCCAGGAGGAGGAGTCCACAGCGGAGCAGCA	70532
Qy	866	CysThrGluGluAlaGluGlyProAlaLysGlyLysAspAla	881
Db	70531	---ACGGTGGCCACTGCGCGGTACACCGGCAACCGGAGTGGCGGCGAG	70475
Qy	882	AlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGly	901
Db	70474	GCGAACAACGCGCGCGCGCGCGAGCTCCCGCAAGGGTGTGGCAACCGGAA	70415
Qy	902	ArgAla	916
Db	70414	ACAGCGACTGAGCGCGGAGCGGCTTAAGGAGCTGACTCTCGCGCAGATGCCAAC	70355
Qy	917	SerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGly	931
Db	70354	GATCCGTGGCGCAAACTGCTTCGAGGCTATCAATGCCGAGGCG	70310

RESULT 51

AC104627/c

LOCUS

AC104627

187921 bp

DNA

linear

INV 22-MAY-2002

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Drosophila melanogaster X BAC RP98-4A11 (Roswell Park Cancer Institute *Drosophila* BAC Library) complete sequence.

AC104627

AC104627.8 GI:21070446

HTG

Drosophila melanogaster (fruit fly)*Drosophila melanogaster*

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 187921)

Muzny, D., Scherer, S., Adams, M. D., Holt, R. A., Evans, C. A., Gocayne, J. D., Tabor, P., Williamson, A., Homs, F. H., Dugan-Rocha, S. D., Sodergren, E. S., Hodgson, A. H., Chen, R. C., Ayele, M., Scott, G. S., Worley, K. W., Amaratunga, P. G., Brandon, R. C., Rogers, Y. A., Baldwin, D., Beeson, K. Y., Brown, M., Buhay, C., Basam, D. A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K., Davenport, L. B., Dietz, S. M., Ding, Y., Dodson, K., Doup, L. E., Draper, H., Emery-Cohen, A., Ferreira, S., Garg, N. D. S., Houck, J., Hostin, D., Howland, T. J., Hume, J., Ibegwam, C., Jalali, M., Kovar, C., Liu, W., Mattei, B., McIntosh, T. C., Morgan, M., Moy, M., Murphy, B., Nelson, K. A., Ndassa, Y., Nguyen, N., Perez, L., Pittman, G. S., Puri, V., Scheeler, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S. M., Xiang, J., Zaveri, J. S., Zhou, J., Zorrilla, S., Smith, H. O., Wheeler, D., Weinstein, G., Gibbs, R. and Venter, J. C.

Direct Submission

Unpublished

2 (bases 1 to 187921)

Worley, K. C., Adams, C., Adio-Oduola, B., Ali-oman, F. R., Allen, C., Alsbrooke, S. L., Amaratunga, H. C., Are, J. R., Banks, T., Barbarella, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N. P., Buhay, C., Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C., Caron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J., Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoque, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Mahehwar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Monabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogulu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodargren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Wu, C., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission

Unpublished

3 (bases 1 to 187921)

Worley, K. C.

Direct Submission

Submitted (15-DEC-2001) Human Genome Sequencing Center, Department

Db 6448 CCACAGGATTCGGTTCGCTACTACTACCTACGCAAGAACCGGAGATTACAAGCAGCTG 6389
Qy 483 ValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln 502
Db 6388 CTTGGAATAATCCCGCAACGACGCGAGTTCGGTAAATCCAGCAAGGCTCAGCGCGCA 6329
Qy 503 Gln-----GlnGlnGln 506
Db 6328 CAGCCGCGAGTCATATTATCGATTTCGATGACCACTGGCGTTATGACGCGATTTCAGCGCGAA 6269
Qy 507 GlnGlnGlnGlnProMetProArgSerSerGln---GluGluLysAspGluLysGluLys 525
Db 6268 CAGCAACAAAGATCCGCGGTCGTTTCATCCCGCTTCTGAGCGTGGAGCGTCCGGAACGT 6209
Qy 526 GluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAspLeu 545
Db 6208 GCCCGCGAGCTGAGCTGTGGCAGAGAAACCGCAGCTGATGCTGCCAA-----6158
Qy 546 LeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaValAla 565
Db 6157 -----GCGGCTGAAAGTCCCGCCGAAAGGCAAGCGCGGCC 6122
Qy 566 SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgLysThrArgSer 585
Db 6121 ACTAAG-----6116
Qy 586 MetAlaAsnGluAlaAsnSerGluAlaIleThrProGlnGlnSerAlaGluLeuAla 605
Db 6115 ---GCTCGAAGCAACGGCT-----6098
Qy 606 SerMetGluLeuAsnGluSerSerArgTrpThrGluGluMetGluThrAlaLysLys 625
Db 6098 -----6098
Qy 626 GlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThr 645
Db 6098 -----6098
Qy 646 ValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle 665
Db 6098 -----6098
Qy 666 LeuGlnGlnHisLysLysLeuLysMetGluLysGluArgAsnAlaArgLysLysLys 685
Db 6097 -----GCGGTGAAAGGTGGCCAAA 6077
Qy 686 AlaProAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluLysMet 705
Db 6076 GCGGCTGCAGCGCGCGCGGCGAGCAACTACCGCCCAACCCGCAACCAACCA 6017
Qy 706 GluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHis 725
Db 6016 TCATCATCAATCATCATCATCATCA-----5987
Qy 726 AlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnSer 745
Db 5986 GCATCATCAGCATCAACAGCATCATCATCAGGCGCAAGTCCAGCGCATTG-----5936
Qy 746 SerAspThrGluSerIleProSerProHisThrGluAlaLysAspThrGlyGlnAsn 765
Db 5935 -----CGAGAAATAGCGCGAGATAAAGCATCCGCGCAAA---ACGGCTAGCGCC 5888
Qy 766 GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro 785
Db 5887 AGCGATAAGATCGAGCAACAGCTGTT-----GGACCA-----5855
Qy 786 ProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAla 805
Db 5854 -----ACGGCAACCGGAAGT-----5840
Qy 806 ProThrProProAlaProProSerProSerAlaProProValProLysGlu 825
Db 5839 CCGACAGCAGCAACTACACCTGCAACGCAACCAACGACCA-----5798

Qy 826 GluLysGluGluGluThrAlaAlaProProValGluGluGluGlyGluGluGlnLysPro 845
Db 5797 -----GAGATCTCAGCGGC-----GGCGAGCCCAAGAGCAA 5765
Qy 846 ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGlu 865
Db 5764 AATCGCGAGGAGGAGGAGCTGCCACAGCGCGAGCAGCA-----5726
Qy 866 CysThrGluGluAlaGluGluGluProAlaLysGlyLysAspAla-----Glu 881
Db 5725 ---ACGTGGCGCACTCGCGGTACACCGGCAACCGGAGCAAGTGGCGCAGTGTGGCGAG 5669
Qy 882 AlaAlaGluAlaThrAlaGluGluAlaLeuLysAlaGluLysLysGluGlyGlySerGly 901
Db 5668 GCGACACGCGCAACGCGCGCCACGCGCACAGCTGCCCGCAAGGTTGGCAACCGGAA 5609
Qy 902 ArgAla-----ThrThrAlaLysSerSerGlyAlaProGlnAspSerAsp 916
Db 5608 ACAGCGACTGAGCGCGGGAACGGCTGCTAAGGAGTGACTCTCGGCCAGATGCCAAC 5549
Qy 917 SerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGly 931
Db 5548 GATCCGCTGCCAAAACCTGCTTGAAGGCTATCAATGCCGAGGC 5504

RESULT 52

AE003490/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

302665 bp DNA linear INV 14-FEB-2003
Drosophila melanogaster chromosome X section 42 of 74 of the
complete sequence.

AE003490 AE002593 AE014298

AE003490.2 GI:22832147

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 302665)

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Anatolides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Yang, M.D., Zhang, Q., He, L.X.,
Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champagne, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L.,
Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D.,
Ballou, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M.,
Beeson, K.Y., Benos, P.V., Bertram, B.P., Bhandari, D., Bolshakov, S.,
Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brotter, P.,
Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A.,
Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B.,
Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I.,
Dietz, S.M., Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S.,
Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C.,
Ferrier, S., Fleischmann, W., Fosler, C., Gabriellian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalili, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kennis, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
Nusskern, D.R., Pacle, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Sidenkamos, I.,
Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R.,
Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A.,
Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodgett,
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,

Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
PUBMED
10731132
2 (bases 1 to 302665)
Celniker,S.E., Adams,M.D., Krommiller,B., Wan,K.H., Holt,R.A.,
Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y.,
Banzon,J., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,K.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,I.J.,
Ibegwam,C., Jallali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunco,J.,
Paclob,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Tyler,D.,
Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.
Sequencing of *Drosophila melanogaster* genome
Unpublished
3 (bases 1 to 302665)
Misra,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L.,
Campbell,K., Hradecky,P., Huang,Y., Kaminker,J.S., Prochnik,S.E.,
Smith,C.D., Tupy,J.L., Bergman,C.M., Berman,B.P., Carlson,J.W., de
Celniker,S.E., Clamp,M.E., Drysdale,R.A., Emmert,D., Frise,E., de
Grey,A.D.N.J., Harris,N.L., Krommiller,B., Marshall,B.,
Millburn,G.H., Richter,J., Russo,S., Searle,S.M.J., Smith,E.,
Shu,S., Smucniak,F., Whitfield,E.J., Ashburner,M., Gelbart,W.M.,
Rubin,G.M., Mungall,C.J. and Lewis,S.E.
Annotation of *Drosophila melanogaster* Genome
Unpublished
4 (bases 1 to 302665)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
5 (bases 1 to 302665)
FlyBase
Direct Submission
Submitted (06-SEP-2002) University of California Berkeley, 539 Life
Sciences Addition, Berkeley, CA 94720, USA
6 (bases 1 to 302665)
FlyBase
Direct Submission
Submitted (23-JAN-2003) University of California Berkeley, 539 Life
Sciences Addition, Berkeley, CA 94720, USA
On Sep 13, 2002 this sequence version replaced gi:7292788.
Location/Qualifiers
1. 302665
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/note="genotype: Y[1]; cn[1] bw[1] sp[1]; Rh6[1]"
3278. 4294
/locus_tag="CG15728"
/note="last curated on Fri Jun 28 07:20:51 PDT 2002"
/map="11B5-11B5"
/db_xref="FLYBASE:FBgn0030409"
join(3278. 3409,3466. 3657,3722. 4294)
/locus_tag="CG15728"
/product="CG15728-RA"
/db_xref="FLYBASE:FBgn0030409"
join(3278. 3409,3466. 3657,3722. 4294)
/locus_tag="CG15728"
/note="CG15728 gene product"
/codon_start=1
/product="CG15728-PA"

protein id="AAF48183.1"
/db_xref="GI:7292789"
/db_xref="FLYBASE:FBgn0030409"
/translation="MAEYNNMSHMVPPQGSLEELRQTMRSRAIRSQCYFIATGNILE
ISGFDLLNQGIQFADKLAAYVCGIMVYLLHEDTSATQVRTFLQKCPYMA
CSESHIHQVCHVIGLIDIQSSSTMLNLILGCRVASPLCTWASVAVCLLWMLDMM
SDLGLHLRPANTLLVAVVKPHIYVSYLHSLHVLVRLISLILGIPNAQQQ
LCQETDVPEDYMKLARDDCSILRWLIAIVDELRLPLMVNNDLGHRLHRLVLESTICE
LWQLLGHVLCVCKSGSGPOCK"
complement(4415. 5643)
/locus_tag="CG15727"
/note="last curated on Fri Jun 28 12:19:58 PDT 2002"
/map="11B5-11B5"
/db_xref="FLYBASE:FBgn0030410"
complement(join(4415. 5455,5520. 5643))
/locus_tag="CG15727"
/product="CG15727-RA"
/db_xref="FLYBASE:FBgn0030410"
complement(join(4600. 5455,5520. 5548))
/locus_tag="CG15727"
/note="CG15727 gene product"
/codon_start=1
/product="CG15727-PA"
/protein id="AAF48184.1"
/db_xref="GI:7292790"
/db_xref="FLYBASE:FBgn0030410"
/translation="MSGKDDRSKERNMKQYNHHKKDQGRSTTSSGSPPTSRSRSM
VDVAVSSDEPLRARGNWQSSGAGRSSLAQNRNMDLTLPSDITDDVEQLDGDAPID
ENARQAGDFQQAQFPGLGGHFTFGSEREWANVAGQTKLHTKAASAYPTLNLT
RLNVGLQITPLYRMDYPASLFTFRAQIAAQEAERAEAYQOCLDKDANGAKSRAP
SAKSNKDVAPAAAEKPIAAASAAEPDELDELLAMTDQLDIGSGTITMPPMPVQAPATPS
SAASKNGVEWLDLSVLDE"
5955. 7164
/locus_tag="CG2540"
/note="last curated on Fri Jun 28 07:21:09 PDT 2002"
/map="11B5-11B5"
/db_xref="FLYBASE:FBgn0030411"
5955. 7164
/locus_tag="CG2540"
/product="CG2540-RA"
/db_xref="FLYBASE:FBgn0030411"
6004. 6939
/locus_tag="CG2540"
/note="CG2540 gene product"
/codon_start=1
/product="CG2540-PA"
/protein id="AAF48185.1"
/db_xref="GI:7292791"
/db_xref="FLYBASE:FBgn0030411"
/translation="MLDFTLELRSSGRADQLFGNTYTFEFANVIEPREHGCKSADQ
DFTQGSMLRTPTDENSENLEPPAATDADVWIFGYSLVKTKDFTPIDRRRGFWGFK
RFPYQHSIDHRGIPERPGRVVTLPGDPAQDRVGVAVYRIASQKAGVLDHLDRKKN
GYRCSLEFHEPTDGAEPQIMVYATQANDSYAGDVQVPCIAQIIFSSAGPSGPN
REYFNLAAMDQLPFGAVDHELELVACVKRYIVDEFPQLINHALHHEISGLREEQ
LAEQAQLEKLLERCKQPGREWLIYAALQAKNKS"
7665. 29389
/gene="tomosyn"
/locus_tag="CG17762"
/note="synonyms: CG11056, CG2205; last curated on Mon Nov
04 07:59:47 PST 2002"
/map="11B6-11B7"
/db_xref="FLYBASE:FBgn0030412"
join(7665. 7858,9963. 10002,11777. 12328,12584. 12681,
12756. 12944,13096. 13230,13306. 13369,13483. 13690,
13752. 13827,14064. 14218,14389. 14570,14638. 15120,
15199. 15341,16707. 16888,27155. 27271,27340. 27462,
27524. 27643,27720. 28169,28286. 28394,28460. 28546,
28614. 28709,28784. 28908,28992. 29389)
/gene="tomosyn"
/locus_tag="CG17762"
/product="CG17762-RR"
/db_xref="FLYBASE:FBgn0030412"
join(9565. 10002,11777. 12328,12584. 12681,12756. 12944,

13096. .13230,13306. .13369,13483. .13690,13752. .13827,
 14064. .14218,14389. .14570,14638. .15120,15199. .15341,
 16707. .16888,27155. .27271,27340. .27462,27524. .27643,
 27720. .28169,28286. .28394,28460. .28546,28614. .28709,
 28784. .28908,28992. .29389)
 /gene="tomosyn"
 /locus_tag="CG17762"
 /product="CG17762-RA"

Alignment Scores:
 Pred. No.: 6.73e-06 Length: 302665
 Score: 743.00 Matches: 237
 Percent Similarity: 41.08% Conservative: 106
 Best Local Similarity: 28.38% Mismatches: 254
 Query Match: 5.62% Indels: 238
 DB: 3 Gaps: 23

US-09-522-753-5 (1-2517) x AE003490 (1-302665)

Qy	144	LysLeuGluProValSerProProSerProProHisThrAspProGluLeuVal	163	Db	88615	GCCTGGAGGATAAAAGATGCGATCCTATCGCGTTATACCGCTCTGTATGACGATGCC	88556
Db	89362	CAGGTGGAGGCCATTCGCGACATTCGCC-----AGCGATAGTTCCGATTGAGGACGC	89309	Qy	403	AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVal	422
Qy	164	ProProArgLeuSerLysGluLeuLeuIleGlnAsnMetAspArgValAspArgGluIle	183	Db	89555	CGCCAGAGCGATGCGCCTATCATATGAAATGGACTTATCGAGGATATGGTGGCAGTT	88496
Db	89308	GGAAGGACCGACGCGCAAGAGGATCTCTCATGCAATCCAAAGAGTTGACATGAGATC	89249	Qy	423	TyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPheArgGlu	442
Qy	184	ThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnGluGlu	203	Db	88495	CATCAACAGCGCTAAGGCCCTCAACATGTGGACCGCGCGAGAGGAGACCTTTAAGGAG	88436
Db	89248	AAATCCGCTCAGACGACAAATGGAACCTTCGCAAAAGAGAAATCCCTCATGGAGGAG	89189	Qy	443	LysPheMetGlnHisProLysAsnPheGlyLeuIleAsnSerPheLeuGluArgLysThr	462
Qy	204	Ala-----AlaLysProProGluProGluLysProValSer-----	215	Db	88435	ARGTATTTGCGATTCGAAACTTTGGAGCCATTTCGCGCTAGTTTGGATCGCAAAATCG	88376
Db	89188	GCCGCCCTGCGCAAGAGCAGAGCGCGCCCAAGAGTTGAACGATAACAATGATCAG	89129	Qy	463	ValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu	482
Qy	216	ProProProIleGluSerLysHisArgSer-----LeuValGlnIleIleTyrAspGlu	233	Db	88375	CCACAGGATTCGTTCTGCTACTACTACCTCAGCAGAGNAGACCAGAGATTACAGACACTG	88316
Db	89128	GAACCGATGTGGAACTACTCTGTCGCGCAGCAATGCTAGCGGAGAGATCTATGCGGCC	89069	Qy	483	ValArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGln	502
Qy	234	AsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGlu	253	Db	88315	CTTCGAAATCCCGCCCAACGCGCAGTTCGCGTAATCCAGCAAGGCTCAGCGCGCA	88256
Db	89068	AATCGGAAGACGGCCCAAGCGCAACATTCCTCGTCGCAAGTATCGAGCGCGCGATGAT	89009	Qy	503	Gln-----GlnGlnGln	506
Qy	254	-----LeuProLeuTyrAsnGlnProSerAspThr	263	Db	88255	CAGCCGCGAGTCATTATCGATTCGATCACCCTGCGCTTATGACGCGATTGCGCGCAA	88196
Db	89008	TCCTCGGTTCCGTCGCGCAGACACCATGTTGCCATTGTATAACGACCATCTCGATGC	88949	Qy	507	GlnGlnGlnGlnProMetProArgSerSerGln---GluGluLysAspGluLysGluLys	525
Qy	264	ArgGlnTyrHisGluAsnIleLysIleAsnGlnAla---MetArgLysLysLeuIleLeu	282	Db	88195	CAGCAACAAGTCCCGCGTCTTCATCCGCGTTCGCTGAGCGTGAAGTGGCGGACGT	88136
Db	88948	GAGGCACTTCCCATGCTTAATACGCGCAGCATCAGAGTCAAAATTCGTGCGCGCTACTTTG	88889	Qy	526	GluLysGluAlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeu	545
Qy	283	TyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyr	302	Db	88135	GCCGCGAGCGTGGCGTGGCAGAGAAAGCGCAGCTGATGTCGCAAA-----	88085
Db	88888	CACATAGAAATTAAGGCGCGCAACCTGGCGGCACAACTCAGGACCTGTGGGAGAGTAC	88829	Qy	546	LeuLysGluLysThrAspAspThrSerGlyLysAspAsnAspGluLysGluAlaValAla	565
Qy	303	AspGlnLeuMetGluAlaLeuGluLysValArgIleGluAsnAsnProArgArg	322	Db	88084	-----GCGGTGAAAGTGCCCGCCGAAAGGCAAGCGCGCC	88049
Db	88828	ACCAAGGATCAGCGGATTGGCAGCAGCATGCGAGGTATGAGGCGCAGTGCACAGAGG	88769	Qy	566	SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSer	585
Qy	323	ArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLys	342	Db	88043	ACTAAG-----	88043
Db	88768	AAAGCAGCGGAGCAAAATTCGCAATCTTCGAAAGGCTCTACGAGGAGTAAAGGAA	88709	Qy	586	MetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAla	605
Qy	343	GlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSer	362	Db	88042	---GCTGTCGAAGCAACGGCT-----	88025
Db	88708	CAGCGGAGGATAAGGAACGTTTC---AATCGCGTTCGTTTCGCGG-----	88667	Qy	606	SerMetGluLeuAsnGluSerSerArgTyrThrGluGluMetGluThrAlaLysLys	625
Qy	363	MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln	382	Db	88025	-----	88025
Db	88666	-----ATTAATCCGAGCGGATCTCGAGAGATCATGATGCGCTTCGAGGACG	88616	Qy	626	GlyLeuGluHisGlyArgAsnTyrSerAlaIleAlaArgMetValGlySerLysThr	645
Qy	383	GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAla	402	Db	88025	-----	88025
Db	88666	-----ATTAATCCGAGCGGATCTCGAGAGATCATGATGCGCTTCGAGGACG	88616	Qy	646	ValSerGlnCysLysAsnPheTyrPheAsnTyrLysArgGlnAsnLeuAspGluIle	665
Qy	383	GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAla	402	Db	88025	-----	88025
Db	88666	-----ATTAATCCGAGCGGATCTCGAGAGATCATGATGCGCTTCGAGGACG	88616	Qy	666	LeuGlnGlnHisLysLysMetGluLysGluArgAsnAlaArgArgLysLysLysLys	685
Qy	383	GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAla	402	Db	88024	-----GCCGTGAAAGAGTGGCCAAA	88004
Db	88666	-----ATTAATCCGAGCGGATCTCGAGAGATCATGATGCGCTTCGAGGACG	88616	Qy	686	AlaProAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluMet	705
Qy	383	GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAla	402	Db	88003	CGCGTTCAGCAGCGCGCGCGCAGCAACTACCGCCCAACACCGCAACACACACACA	87944
Db	88666	-----ATTAATCCGAGCGGATCTCGAGAGATCATGATGCGCTTCGAGGACG	88616	Qy	706	GluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHis	725
Qy	383	GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAla	402	Db	87943	TCATCATCAACATCATCATCATCATCA-----	87914
Db	88666	-----ATTAATCCGAGCGGATCTCGAGAGATCATGATGCGCTTCGAGGACG	88616	Qy	726	AlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSer	745
Qy	383	GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTyrAspAla	402	Db	87913	GCATCATCAGCATCAACAGCATCATCATCGACGCAAGTCCAGCGCATTCG-----	87863
Db	88666	-----ATTAATCCGAGCGGATCTCGAGAGATCATGATGCGCTTCGAGGACG	88616				


```
QY 746 SerAspThrGluSerIleProSerProHisThrGluAlaAlaIysAspThrGlyGlnAsn 765
Db 87862 -----GCAGGATAGCGGAGATAAAACGATGCGCGCAA---ACGCTAGCGCC 87815
QY 766 GlyProIysProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro 785
Db 87814 AGCGATAAGTAATCAGCAACAGCTGGT-----GGACA----- 87782
QY 786 ProArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAla 805
Db 87781 -----ACGCGAACGGGAAC 87767
QY 806 ProThrProProAlaProProSerProSerAlaProProProValProIysGlu 825
Db 87766 CCGACAGCAGCAACTACCTGCAACAGCAACAGCACCA 87725
QY 826 GluIysGluGluThrAlaAlaProProValGluGluGluGluGlnIysPro 845
Db 87724 -----GAGATCTCAGCGGGC-----GGCAGGCCAGACCAA 87692
QY 846 ProAlaAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValIysSerGlu 865
Db 87691 AATGCCGAGGAGGCGAGCTGCCACAGCGGAGCAGCA----- 87653
QY 866 CysThrGluGluAlaGluGluGlyProAlaIysGlyLysAspAla-----Glu 881
Db 87652 ---ACGCTGGCCACTGCGGTACACCGGCAACCGGACCAAGTCGGCGCAGTGTGCGGAG 87596
QY 882 AlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGlyGlySerGly 901
Db 87595 CGCACACGGCAACGGCGGCCAGCGCCACAGCTGCCCGCCAGGGTGTGGCAACCGGAA 87536
QY 902 ArgAla-----ThrThrAlaLysSerSerGlyAlaProGlnAspSerAsp 916
Db 87535 ACAGCGACTGAGCGAGCGGGAACGGCTGCTAAGGAGCTGACTCTCGGCCAGATGCCAAC 87476
QY 917 SerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGly 931
Db 87475 GATCCGCTGGCCAAACTGCTTCGAGGCTATCAATGCCGAGGCG 87431

RESULT 53
LOCUS AC127533 165791 bp DNA linear PRI 18-JAN-2003
DEFINITION Homo sapiens chromosome 15, clone RP11-1363020, complete sequence.
ACCESSION AC127533
VERSION AC127533.9 Gi:27802039
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165791)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N.,
Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Canarata,J.,
Chang,J., Choepe,Y., Collymore,A., Cooke,P., Corum,B.,
DeArrellano,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P.,
FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A.,
Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
2 (bases 1 to 165791)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N.,
Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Canarata,J.,
Chang,J., Choepe,Y., Collymore,A., Cooke,P., Corum,B.,
DeArrellano,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P.,
FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A.,
Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
```

```
TITLE
JOURNAL
REFERENCE
AUTHORS
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Submitted (17-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 165791)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,
Canarata,J., Chang,J., Choepe,Y., Collymore,A., Cooke,P., Corum,B.,
DeArrellano,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P.,
FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Mathews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 165791)
Birren,B., Nussbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N.,
Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Canarata,J.,
Chang,J., Choepe,Y., Collymore,A., Cooke,P., Corum,B.,
DeArrellano,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P.,
FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A.,
Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (18-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 18, 2003 this sequence version replaced gi:27369453.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27720
Center clone name: 1363_O_20
----- Location/Qualifiers
1. 165791
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-1363020"
/clone_lib="RPCI-11 Male BAC segment 5"
1. 219
/repeat_region
221..242
/repeat_region
```


[illegible]

QY	517	nGIuGLyLysAspGIuLySgLIuLySgluLaGluLySgluGLuLySProGl	537
DB	149831	AGCCAGCCAAAGCAGCCCAAGCCAGCCAGCCAGCCAAAGCAGCCAGCCAAAGCCAGC	149890
QY	537	uValGIuAAsnAspLySGIuAsPLeuLeuLySgluLySThrAspAspThrSerGIuAAs	557
DB	149891	CAAGACAGCCAGCCAGCCAAAGCCAGCCAGCCCAAGCCAGCCAAAGCAGCCAGCCAGC	149942
QY	557	pAnAspGIuLySgluLaVaAlaSerLySglYArgLySthrAlaAsn-SerGIuGLyA	577
DB	149943	-CAAGACAGCCAGCCAGCCAAAGCCGGCCAAAGCCAGCCAAAGCAGCCAGCCAAAGCCA	150001
QY	577	rGArgLySGlyArGIleThrArgSerMetAlaAsnGluAlaAsn-----SerGIuGLuA	595
DB	150002	GCCAGCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150061
QY	595	la-IleThrProGlnSerAlaGluLeuAlaSerMetGluLeuAAsnGluSerSerArg	614
DB	150062	CCRAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150118
QY	615	TripThrGIuGLuMetGIuThrAlaLySglYLeuLeuGLuHiisGIyArgAsnTrp	634
DB	150119	TGGCCAAAGCACGCCAGCCACTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150178
QY	635	SerAla-----Ile-AlaArgMetValGIYSerLySthrValSerGlnCysLySAsnPh	652
DB	150179	TCAGCCAAAGCCACTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCCTT	150238
QY	652	eTy-rPheAsnTrx-----LysLySArgGInAsnLeuAspGluLeuGLuGlnHisLy	670
DB	150239	CGAGCCAGCCAGCCGGAGAAGCGTGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCTAA	150298
QY	670	sLeuLySmetGIuLySgluArgAsnAlaArgAgLySLySLySLySLySLySalalProAlaAla	690
DB	150299	GCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150358
QY	690	aSerGIuGLuAlaAlaPheProValValGIuAspGIuGLuMetGluAlaSerGIyVa	710
DB	150359	AGCCAGACAGCCAGCCATCCAGCCAGACAGACAGCCAAAGCAAAGAACAGCCAGCCAGC	150418
QY	710	lSerGIyAsnGIuGLuMetValGIuGLuAlaGluAlaLeuHisaSerGIyAsnGI	730
DB	150419	CAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150478
QY	730	uValProArgGIyGluCysSerGIyProAlaThrValAsnAsnSerSerAspThrGluSe	750
DB	150479	CAGCCACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	150526
QY	750	rIleProSerProHisThrGluAlaAlaLyAspThrGlyGlnAsnGlyProLySProPr	770
DB	150527	AGACAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150586
QY	770	oAlaThrLeuGIyAlaAspGlyProProGlyPro-----ProThrProProAr	787
DB	150587	A-----GACAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150631
QY	787	GAArgThrSerArgAlaProileluproThr--ProAlaSerGluAlaThrGIyAlaPro	806
DB	150632	CAAGCCAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150691
QY	807	ThrPro-----ProProAlaProProSerProSerAlaProProValValProLyS	824
DB	150692	CAGCCAGGCAGGCAACCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150751
QY	825	GIuGLuLySgluGLuThrAlaAlaPro-----	835
DB	150752	CAAGAGAGCCAGCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150811
QY	836	ProValGIuGLuGLyGIuGLuGlnLySProProAla-----AlaGIuGLuLeuAlaVal	853
DB	150812	CCAGCCAGTACCCAGTAGTCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	150871

Qy 854 AspThrGlyysAlaGluGluProValIysSerGluCysThrGluGluAlaGluGluGly 873
Db 150872 CAAGCCAAAGCAGCAGCAACAGCCA-----AGCCAGCCAGCCAGCCAGCCAGCCAGCC 150925
Qy 874 ProAlaIysGlyLysAspAlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAla 893
Db 150926 AGCCAAAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 150985
Qy 894 GluIysLysGluGlyGlySerGlyArgAlaThrThrAlaIysSerSerGlyAlaProGln 913
Db 150986 AGCCAGCCA-----AGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA 151030
Qy 914 AspSerAspSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAsp 933
Db 151031 GACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151087
Qy 934 LysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAla 953
Db 151088 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151141
Qy 954 AsnAlaSerProGlnLysProLeu---AspLeuLysGlnLeuLysGlnArgAlaAlaAla 972
Db 151142 CGAAGCCAGCCAAATCCACCAGCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151201
Qy 973 IleProProle---GlnValThrLysValHisGluProProArgGluAspAlaAlaPro 991
Db 151202 ACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151261
Qy 992 ThrLysProAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAla 1011
Db 151262 GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151321
Qy 1012 ProGln-----GlnProGlySerSerProArgGlyLysSerArgSerProAlaPro--- 1028
Db 151322 AGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151381
Qy 1029 -----ProAlaAspLysGluAlaPheAlaAlaGluAla 1039
Db 151382 GCCAGCCATCAAGCTAGGCAAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151441
Qy 1040 GlnLysLeuProGlyAspProProCysTrpThrSerGlyLeuProPheProValProPro 1059
Db 151442 GCCAAG---CCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151495
Qy 1060 ArgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyAlaPro 1079
Db 151496 -----AGCCACCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151540
Qy 1080 ProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro 1099
Db 151541 CCGTCCAAAGCCA-----GCCAAGCCA-----CCAGCCAG 151570
Qy 1100 ProThr-----IleSerAsnProProProLeuIleSerSerAlaLysHisProSerVal 1117
Db 151571 CCAACGAGCTAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151630
Qy 1118 LeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTy 1137
Db 151631 ACAGCCAAAG-----CCAGCC 151645
Qy 1138 SerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAsp 1157
Db 151646 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151704
Qy 1158 ProLysLysLeuAlaProPheSerGlyValLysGln----- 1169
Db 151705 ACCCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151764
Qy 1170 -----GluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyVal 1186
Db 151765 AGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151824
Qy 1187 ProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValPro----- 1204

Db 151825 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151884
Qy 1205 ---GlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIle 1223
Db 151885 CGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151944
Qy 1224 ThrTyArgGlySerIleThrHisGlyThrProAlaAspValLeuTyLysGlyThrIle 1243
Db 151945 CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 151980
Qy 1244 ThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeu 1263
Db 151981 -----AGCCGCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152004
Qy 1264 ProLysGlyHisValIleTyThrGluGlyLysLysGlyHisValLeuSerTyThrGluGly 1283
Db 152005 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152061
Qy 1284 MetSerValThrGlnCysSerLysGlu-AspGlyArgSerSerSerGlyPro----- 1300
Db 152062 CTACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152121
Qy 1301 -----ProHisGluThrAlaAlaProLysArgThrTyAspMetMetGluGlyAr 1317
Db 152122 AGCCAAAGACACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152181
Qy 1317 gValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPr 1337
Db 152182 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152226
Qy 1337 oGluArgHisSerProHisLysGluGlnHisIleArgGlySerIleThrGln 1357
Db 152227 CAAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152278
Qy 1357 nGlyIleProArgSerTyThrValGluAlaGlnGluAspTyThrArgArgGluAlaLysLe 1377
Db 152279 -----CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152322
Qy 1377 uLeuLysArgGluGlyThrPro-----ProProProProSerA 1391
Db 152323 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152382
Qy 1391 rgAspLeuThrGluAlaTyLys---ThrGlnAlaLeuGlyProLeuLysLeuLysProA 1410
Db 152383 ACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152432
Qy 1410 laHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProA 1430
Db 152433 -----CCAGTCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152466
Qy 1430 rgGluGluLeuArgHisThrProGluLeuProLeuAla-----ProArgProLeuL 1447
Db 152467 GCCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152526
Qy 1447 ysGluGlySerIleThrGln-----GlyThrProLeuLysTyAspT 1461
Db 152527 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152586
Qy 1461 hrGlyAlaSer-----ThrThrGlySerLysLysHisA 1472
Db 152587 GCCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152646
Qy 1472 spValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAsp- 1491
Db 152647 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152705
Qy 1492 --ValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyThrGluGluSerLeuLysSerA 1511
Db 152706 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 152765
Qy 1511 rgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValP 1531


```
gene
1661..7081
/ gene="amyX"
CDS
1661..7081
/ gene="amyX"
/ note="SAV5982"
/ codon_start=1
/ transl_table=11
/ product="putative pullulanase"
/ protein_id="BAC73694.1"
/ db_xref="GI:29609645"
/ translation="MDLIPKGSAPRTRTHRSRTAAVAAMAAVLAVSLAOPLEARAAT
PPAPSDAKLAEPARHDAFREQFYFMPDRFANGDTSDNGKGLTGSRLSTGYDPTDK
GFYQGGDLKGLTRKLDYIKGLGTTISWLAPIFNKQVQGTGKQDASAGHYGWIPTFK
VDPHFGTKDLETLISKAHAKMKVFEDVITNHTADVDEEKSYGILSKGAPPLYLK
DGRPDGADYDGRKFPFPAVDGDSFRTPAVAARKNAKVPWLNQDPTWKHVNTEFWTO
AGSSHTGDSGLDLDLWTERPEYVGRMEKIYEKWRDFGIDGFRIDTFPHVNAEVSQ
WATALDAYAKRGKDDYFYFMEGEVYSADTSVTSYVYVQGLDLSLDPFDQDAARSASQ
GSAKSLASVFGDDYKYTTDKANAYEQVTFILGNDHMRIGIFLNQDNPKATDAELLRK
DRLANELMFLSRGNPVVYVYDGEQFTGSGDDKDAQRTMFASKVADYLDDEIGTDKGH
ASDAYDTSAPLYKEIALSKLRDNPALADGIQTERYAADGAGVAFSTDAITGTEY
VAVNNAKASATFATGADTAFTGHTGDDVLKSDADKKITVTPAGAAVVLKAG
RGTTPAAKPSLTAKPADAGATFVLSADVDGGLNARVFAOVGNWRTILGSADHA
PYRTQIKGDVPAGTALRYKAVVIDAAGTASATASTTGTTPAATPTASSRDVAI
VHYKRPDGYDMLYAWGLDAGESTTWPAGHDVFGRDAYGAFVVKLKGDSATWF
LVIDKQDGVDSADRTIDTWSKPLEPVRTDAYGAFVEPLTDGATLSISYIHKDEKDL
DLCGTGLVHTGATPTDMSKPLEPVRTDAYGAFVEPLTDGATLSISYIHKDEKDL
SADSLDLTADGHEWMLNGQENHLLPQAPAGSAAALDLTTSKAWIDRNTVAWNGSDA
AASQTLSSRDGSTAVKDGSLTSDDERWLRLSKTSLTDQAQAAPHLKSTYAWSDVR
SVADRTVPADKRGTYLFTDKNSDGRHLQLAESGTSYVHLLPAPDIATIAEKSQGO
DDEARVRLAAGVAVRAANGAVLAATGVLQGLVDLDLYDAKDLGPTFRGCHPTL
AVMAPTAQSVLELDGAVHVRKNNATGMSVTPGASWKGKPYRVVVKWAPVTRKVY
TNKVTDPYSVALTIDTSERSLVLDLDDLSLAPSGWSLKKPKAVPLRDAEIOELHIDF
QATCDCLASVAAADSEKQOECLTVAAXKDAYNMGVDYPHYTVPESYATDANGTRRIVE
PRMWKLSNOODGLVAMDMVNVNHTAAAGAGTSVLDRIVPGYTORLLADGVALTCC
ANTATENAMGKLVDLSLTVWAKEYKVDGFRFDMGHQPKANILAVRKALDALTVAKD
GVDSKKIILIEGCHNFEVADDAFVQATOKNMGATGIATFSRDARDVARGGPFAD
PVDQFGGLGTYDPSNSDANGTPAEQKARLLHQQDLIKVGLSGNLAKYRPTDSSGKEV
TGSVDYNGTAGADAPGADALAYADAHNDSYDLSYDLYKLPKGTAGDRARQVILAM
ATAALAQPSLSQASGSLLSKSLDRNSYDSDGNFNAIHWNCQDNGFGRLPMAADN
KSKWPYATPLTSLVKVGCQDTEGTSAGYQDLIRITTEPDSLSIAGVQVQSKLTFPLS
GKDETTPGVTIMKLGDLVVVFNAIPDQOEQTVAALAGKDYALHPVQAAGADPIVKSASY
TAKSGMFAVPGRTVAIFSQVAR"
7152..8756
/ note="SAV5983"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="BAC73695.1"
/ db_xref="GI:29609646"
/ translation="MDANGRTDSTVMVVDVPTSRVAMGAVLRAGHSVVPVASAGE
ALVELDLVLRAGSLPVDVALVDGFLPDMGSEFELCRRLKSQPTPALPVVHPSAALAPG
DRCCQLDAGGDAYLSVPAEPEEIQAVVRAAVRGARLRTGAALARRLITLSETIVAVQ
NAARCTELAHAAQAOTARLTGDFAAVFLVIGEDGELYPLGSLRRARTSLPGGAHEVA
LRIGCTTAGHAGVRSIVVPAWLPAGFLPGVSESDARVLAQTOGKPGVACLTAPFAH
PAQETGALVARLAQATAALAAPFLMYQVVERHVALTLOHSFLPKQPADLPGMVEAVRVV
PASRQTEIGGPFYATLTSRGLVITGVGVGHSLSDAATVMVEIRHALLRVCVDPDPA
VLVTRIDRMUQHYHEPATVCLALVDPSDHSVSIANAGHIPPILLVHDDGTARVPGA
GPLGLGLRPERPATEEFITPRHRLMLVTDGLIETRTGDLTSLVSMQELRMAAADGSPGVG
ALCDALLDFCGRDRDDELLALLRLG"
8788..9144
/ note="SAV5984"
/ codon_start=1
/ transl_table=11
/ product="putative isomerase"
/ protein_id="BAC73696.1"
```

```
/db_xref="GI:29609647"
/ translation="MPOITVDVYSAEIAPGDFDRPGFARALHEEVVVRTAAAKPEACKTOF
RGGEDTVDGAGEGHEIIVHTLIGLLAGSDETKARLTEAALELLRQVVKPDGGLALHA
SAEVRTDIPGSRKFPDA"
complement(9148..9849)
/ note="SAV5985"
/ codon_start=1
/ transl_table=11
/ product="putative TetR-family transcriptional regulator"
/ protein_id="BAC73697.1"
/ db_xref="GI:29609648"
/ translation="MTTGVRRMRMGVEERRQQLIGVALDLFSRRSPDEVSIDEIASAAG
ISRLVYHYFPGKLSLYEAAQLQASDLAGRFVEPRGPGIARLLRWRRFFDFVEEH
GGPSALMRGGPAGVSSSTNALVDGQAAVQILSHLKVTEPPVRLVLEIWRISLV
ESTALLDGRRIARGELEVLVHDFALAAIAISAYDDELAALLRDVLRDAPDGPFS
DLVRLIALTSPPTA"
complement(9928..11844)
/ note="SAV5986"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="BAC73698.1"
/ db_xref="GI:29609649"
/ translation="MQSDVAPRPTPASGPVVDVEQAAALVVEHPVRLVRLAYLVLP
SLGRNRLVTAHALTQRALPRGRTSAVIPAQKGRDGDGVALVRLHVVRTALEAGL
KLSEDPVRKRSQLPPLPQVWGLRFLPRSGGADELALDQRLSALSGPAAAHVLRGLE
KLSERVRVLDAGVADPAIAIEAAKVPAPDALLDPSFDPSCSLHARPTDLMRRO
HMKALAAAAYVCGALLGLPGGWA PDGAAAPSAENPAARALDARLTKTAAATA
WKTARTDPSVWPARGDTGTALRLRALVWAPGEAVRSATPGTSGAPAGPQL
LYAGDSAVHVVLLYDGLRIARYAEPKDTGMAALDFARVDGASGAEEALVLRADG
NVRYLTAPWVTGAERDLMKPSAGAMDNLTDGVAAPLAGSAQRTGACTSNWALQITD
SGSTRLTLDGELI PAHLTTGPSPAPREASRTAALHSPAFCASLASVRSQGVRSVA
WQFAOQALPDGSGSAGVCTRADTWGGGTRVLAQFTPGGAYGAFAAKAEDSPACGA
RDPHVLAGVLMKSGTGAVLLAAGSSDTASIRATGCGIDGAAEGNLLTVRTERGARGL
KGVLDGRPISGLR"
complement(12002..12514)
/ note="SAV5987"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="BAC73699.1"
/ db_xref="GI:29609650"
/ translation="MGVPTTENWDERVAAAANAFTDDYAEKDAADFRAVVDALVAELPA
DSPFGPFEQACAWDSTGSHDKAWPLVIREALAGLSGKGRRAKIQLSSLRNIGQAE
GVKLLTPELEPSDELDDAVRACTALCLADGREREGUSLVILGALAPHLPRYQRSWAN
YARLLVEPDD"
complement(12571..15687)
/ note="SAV5988"
/ codon_start=1
/ transl_table=11
/ product="putative alpha-N-acetylglucosaminidase"
```

Alignment Scores:			
Pred. No.:	1-226-05	Length:	299925
Score:	730.00	Matches:	703
Percent Similarity:	30.06%	Conservative:	291
Best Local Similarity:	21.26%	Mismatches:	1119
Query Match:	5.52%	Indels:	1201
DB:	1	Gaps:	148
US-09-522-753-5 (1-2517) x AP005045 (1-299925)			
Qy	15	ThrGluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaAatThrHis	34
Db	259302	ACAGACCAGCACATCCCATCTCTGCTCACCGGCTTCGGCTCGGACCCGAACCTCAC	259361
Qy	35	ThrAspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSer	54
Db	259362	-----CACACCGC-----	259370
Qy	55	ProGlySerIleLeuGlnProGlnArgArgProSerLeuLeuSerGluPheGlnPro	74
Db	259371	-----CTTCGCACTCGCGCTCGCGA-----CCT	259394

Qy 2300 snGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIle- 2319
Db 267138 AGCGAATACCGCTTCGG---ACGCTTCGGGACC-----CTTCGGCGCCCGCCCTTC 267188
Qy 2320 -----Thrg 2321
Db 267189 ACCTCGCGGCTCGCGCGCGGAGCGCGACCGCGCGGTTCACGGGACCCCGCGCCCG 267248
Qy 2321 lyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetG 2341
Db 267249 GTACGGCGGACTTCTTCAGAACTTCGACGGTTCAACGGCATGTCTCGTACTCTTCGG 267308
Qy 2341 lyLeuGluAlaIleAlaArgIleAlaLeuMetGlyIleTyr- 2354
Db 267309 GAAAGGCGAATGATCGCGGTAGTCTTGCGCCACCGACCGACCGCGCGCGCGCGCGCG 267368
Qy 2355 -----AspGlnTrpGlu- 2358
Db 267369 AGTTTGCATCTGATGACAGTACGACCCACCGCGCGCGCGCGCGCGCGCGCGCGCG 267428
Qy 2359 -----GluSerProp 2362
Db 267429 TATGACCCCGCGGAGGAGAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 267488
Qy 2362 roLeuSerAlaAsnAlaPheAsnProLeuAsn-----A 2373
Db 267489 CCGCATCGCGTCCGCGCGGAGGAGCCACCTCGAAGTCGACCGCGCGCGCGCGCGCGCG 267548
Qy 2373 laSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHis 2393
Db 267549 CATCACCGTCAGCGCGCGCGCGCA---CCGACAGCGCGCGCGCGCGCGCGCGCGCG 267590
Qy 2393 hrLeuThrSerProGlyGly-----GlyGlyLysAlaLysAlaValSerGlyArgp 2409
Db 267591 CACGGACCTCACCG 267650
Qy 2409 roSerSer-----ArgLysAlaLysSerPro----- 2417
Db 267651 CGTCGACGACTGGAAACCGCGCGCTTCTCCATCACCCCTGAACTACCAACACGCGCG 267710
Qy 2418 -----AlaProGlyLeuAlaSerGlyAspArgPro----- 2428
Db 267711 CGGCGAAGCATCGCGCGCTTACCAACCGCGCGGTCCATCACCGCGCGCGCGCGCG 267770
Qy 2429 -----SerValSerSerValH 2434
Db 267771 CACGGACTGGGGAACCTACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 267830
Qy 2434 isSerGluGlyAspCysAsnArgThrProLeuThrAsnArgValTrpGluAsp---- 2452
Db 267831 CCCTGTGTGGCCCTGCG 267890
Qy 2453 -----ArgProSerSerAla---GlySerThrProPheProTyrAsnProLeuIleMetA 2470
Db 267891 GCGCGAAGCGCTCG 267950
Qy 2470 rgLeuGlnAlaGlyValMetAlaSerPro-ProProGlyLeuProAlaGlySerGly 2489
Db 267951 CCTGTGTAACCTGCG 268004
Qy 2490 ProLeuAlaGlyProHisAlaTrpAspGluGluProLysPro 2504
Db 268005 CCG 268048

RESULT 55
AR166425/c
LOCUS
DEFINITION Sequence 3 from patent US 6280999.
ACCESSION AR166425
VERSION AR166425.1 GI:16241741
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 33529)
AUTHORS Gustafsson,C., Betlach,M.C., Ashley,G., Julien,B. and Ziermann,R.
TITLE Sorangium polyketide synthases and encoding DNA therefor
JOURNAL Patent: US 6280999-A 3 28-AUG-2001;
FEATURES
Location/Qualifiers
1..33529
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 2,47e-06 Length: 33529
Score: 726.50 Matches: 674
Percent Similarity: 32.36% Conservative: 314
Best Local Similarity: 22.08% Mismatches: 1199
Query Match: 5.50% Indels: 873
DB: 6 Gaps: 137

US-09-522-753-5 (1-2517) x AR166425 (1-33529)

Qy 17 ProArgTyrProPro---HisSerLeuSerTyrProValGlnIleAlaAArgThrHisThr 35
Db 22686 CCGAGGCTCCACGACACCATCGCTCCCGCGCGCTCGCACCTTGA-----TCAGCC 22633
Qy 36 AspValGlyLeuLeuGluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerPro 55
Db 22632 GACCTCGGG-----CGCACACTAGTCTTCTCGCTCGCGCA 22597
Qy 56 GlySerIleLeuGlnProGlnArgArgProSerLeuLeuSerGluPheGlnProGly 75
Db 22596 GGCTCAGCACACCGCGCACATCGCAGCGCTCTCTATCGAGT---GCCCGACGA 22540
Qy 76 AsnGluArgSerGln-----GluLeuHis----- 83
Db 22539 GCGAGCGGCTCAGCCCCCAGCACTCCACTGAGGTACAGCGCCACTTCCAGCGCA 22480
Qy 84 -----LeuArgPro---GluSerHisSerTyrLeuProGluLeuGly 96
Db 22479 CAGCGCGGCTGGGTACTCGCTCGCTCAGCAACGCTCTCTCTCGTCTCGTCCGCGG 22420
Qy 97 LysSer-----GluMetGluPheIleGlySerLysArgProArgLeuLeuLeu 113
Db 22419 CGCGAACACCACTCTCTCAACCCAGCGTCCAGATCGCTCCAGCGCTCGCACACCTC 22360
Qy 114 ProAsp-----ProLeuLeuArgProSerPro 122
Db 22359 GTCGAACCGCGCACGGAACACGGGTACACTTCGTAAAGCTCTTCCCATCCCGAGCG 22300
Qy 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db 22299 CTGGTGCCTCG-----CCCGGTGAACAGCAC 22273
Qy 143 GlyLysLeuGluProValSerProProSerProHisThrAspProGluLeuGluLeu 162
Db 22272 CGCAAGCTT---CCCGCTCGCTTCGCGCGTCCGCTCAG-----CACCGCGCGATC 22225
Qy 163 ValProProArgLeuSerLysGluLeuIleGlnAsnMetAspArgValAspArgGlu 182
Db 22224 CGCGCGCGCGCGACGACGCGCGGAGACCTCCACAGCTCCCGCAGCGCTCGCGCGCA 22165
Qy 183 IleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGlu 202
Db 22164 CACGAGCGCGGACTCGGAAGTGGCTCGCGTCCGCGCGCGCGCT-----GCG 22117
Qy 203 GluAlaAlaLysProProGluProGlu-----LysProValSerProProProle 219
Db 22116 CACCGCTCGCGACACCGCGCGCTCTCTCCAGCGCACTTCGCCCGCGCGCGCGCG 22057
Qy 220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsn-----ArgLysLys 237
Db 22056 CTGCGCATTCACCGCGCGCTCTCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 21997

QY	238	AlaGluAlaAlaHiaArgIleLeuGluGlyLeuGly----	249	538	lGluAsnAspLysGluAspLeuLeuLys---GluLysThrAsp---AspThrSerGlyGlu	556
DB	21996	CGCTCGGCTCGACAGAGCTCGCGCGGCTCGCGCGGCTTCTTCGAGGATGATG	21937	21079	CGACAGTATGAAGAGACACCGTATGTCGAGCCCATCGAACCCAGGTACACTCCGGTGGCC	21020
QY	249	-----	249	556	uAspAsnAspGluLysGluAlaValAlaSerLys-----	567
DB	21936	CGCGTTTCGCTCCGCTGATCCCGAACGACGACGCGCGGCGACGACGCGACCGCGGTTGCG	21877	21019	TCCGCTCAGCGCGGCGGCGACGCGCGGCGCTCCAATGCCTCCACGCGCTCCCA	20960
QY	250	---ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgLysHisGlu	268	568	-----GlyArgLysThrAlaAsnSerGlu	575
DB	21876	CGGCCACGAGCGGCTCTTGACGAGCGACGACCGGCT-----TCCTTCCACCC	21826	20959	GCACACCGCATGCTCGGATCCATCGACTGCGCTTCGCGCGCGGATA-TCCCGAAGAAG	20901
QY	269	AsnIleLysIleAsn-----GlnAlaMetArgLysLysLeuIleLeuTyrPhe	284	575	nGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAla	595
DB	21825	AATGTCGGGCTCGGCTGCTCGCGTGCAGCGCTCTTCGACGACCTCGTGTGTCATCGA	21766	20900	TTCGCTTCGAAACAGGTCGATGTCGCCAAACACCACTTCGCGCGCTGCTTCGCT	20841
QY	285	LysArgArg---AsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAsp	303	595	allieThrProGlnGlnSer-----AlaGluLeuAlaSerMetGlu	608
DB	21765	CAGACCATCTTGATCACCCCGC-----	21742	20840	CGGCTCCGATCAGGTCGTAGACTCGAACCTCGAACCGCTCGGAGCGCTTCG	20781
QY	304	GlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsnProArgArg	323	608	uLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLe	628
DB	21741	-----CACGCGCGCGCGC	21727	20780	ATCGCTCTTCCCTCGGCCAAGACGCGCATGCTGTCGGGCTCGCAAGCGCTCCC	20721
QY	324	AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluLysArgLys	342	628	uGluHisGly-ArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerG	648
DB	21726	CTCGCATGTCTCCAGGTTCTGA-----CTTCGACGATCC	21694	20720	GGCAGCGCGCACGCCATCGACGATCGGTCGTCTTCTTCCTGCACGACGACGC	20661
QY	343	-----GlnArgGluLeuGlnArgMetGlnSerArgValGlyGlnArgLysSerGly	360	648	lnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeu-----	662
DB	21693	CAGGTACAGCGAGCTCGGGGCTACGCTCCGGTCAAACACTCGGCAGCGCTCGGC	21634	20660	ACCGCGCGCGCTTCTTCACTCGACGCCCGCAGAACCCCTGCTTCACGACGACTCT	20601
QY	361	LeuSer---MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu	379	663	-----AspGluIleLeuGlnHisLysLeuLysM	673
DB	21633	CTCGATCGGCTCTCCAAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	21574	20600	CGGATCCCGCGCGCTCGGCTAGTCAAAGACACGAGCTCGCCGCGCAAGGACGCT	20541
QY	380	SerGluGlnGlnAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeu	399	673	etGluLysGluArgAsnAlaArgLysLysLysLys-----	685
DB	21573	GGGCGACAGACCCACA-----CGACGACAGCGCTCGCGGATCACCGCTGCTG	21526	20540	CGGACGTCGAGCGGTTGCGCAGCGCCCGCATCAGCGAGTCGAGCCCCAGCTCTCT	20481
QY	400	TyrAspAlaAspGlnGlnArg-IleLysPheIleAsnMetAsnGlyLeuMetAlaAspPr	419	686	-----AlaProAlaAlaLaserGluGluAlaAlaPheProp	698
DB	21525	GGCAGGCGCTTCGGCGCGTCAGACCTGGCTGCGACCGTCTCTGTTCCCGCGAGCC	21466	20480	GCACCTGATCGCTGCGACAGCTCGCGCGCTGAGCCCTGCCCGCGCGCACCTCGC	20421
QY	419	oMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGlnGlnLysGluTh	439	698	roValValGluAspGluGluMetGluAlaSerGlyValSerGlyAsnGluGluGluMetV	718
DB	21465	ACGGATCACCCCGACGACGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGC	21409	20420	CCGCGACGCTCGACGCGCATTCAGCGCTCGCGCTCGCGCGCGCGAGCGCGCT	20361
QY	439	rPheArgGluLysPheMetGlnHisProLysAsnPheGly---LeuIleAlaSerPheLe	458	718	alGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerG	738
DB	21408	-----CAACAGCATCCGCAACCTCGCGCGCGCGCGCGCTCGCGCT	21365	20360	CGCGAGCGCGCGGCTGCTGCTGCTGCGAGGACGCTTCGCGAAG-----CTCG	20307
QY	458	uGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAs	478	738	lyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSer-ProHisThrGlu	757
DB	21364	GCACGAGAAGCTCTTGACCGACCGCTCGCGGACATCCCTT-----GAGAC	21317	20306	GCGCAACAGCGCACGAAACAGCGCGCGGCTCGCGCTGAGCTCCAGCGCCAGCTGCA	20247
QY	478	nTyrLysSerLeuValArgArg-----SerTyrArgArgArgGlyLysSerGlnGlnG	496	758	AlaAlaLysAspThrGly-----	763
DB	21316	GGCTGAATCTCCAGAACGCGGGGGTGTCTATCATCCGCTACCCCGCGCGCGCGCA	21257	20246	GCTGCGGATCCAGGTGCACTGGGACCGGCTCGTTCAGGCGGTGAGAGCGCGCGCT	20187
QY	496	n-----GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	509	764	-----GlnAsnGlyProLysProAlaThrLeu-----	773
DB	21256	GGTCGATTCGCGCGGCGCGCGCTGACGCCAGGTGACGACACCCAGCGACGCG	21197	20186	CGACGAGCGAGCGGCTCTTCCACCGGATCGCGACGACCCCATTTGGCTCATGCGCG	20127
QY	510	-----GlnProMetProArgSerSerGlnG	518	773	-----	773
DB	21196	AGCAGCGGTGTCACCGTATCGCTGGCGCTGACGCCCCAG-----CGGTAGGCCACCC	21140	20126	ACAGCTCGGCTCGCCCCAGGTGCTGTATCCCAACGCGCGCTGTCTCCAGAAGCCCC	20067
QY	518	uGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluLysProGluVal	538	774	-----Gly	774
DB	21139	GGCCCGAGAGGATCTCGCGCGCTCCGATGCCCCGGTACCGGTCCTTCAGTCTTCAGAT	21080	20066	ACGCAGCGCTCGTCCGCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	20007
QY	538	lGluAsnAspLysGluAspLeuLeuLys---GluLysThrAsp---AspThrSerGlyGlu	556	775	AlaAspGlyProPro-----ProGlyProPro-----ThrProProArg	787

Db 2006 ACGTGTTCCGCGCGGTAGTGTCTCTGGCGCGCGCGCAAAAGTACCGCGCGCGACG 19947
QY 788 ThrArgSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThr 807
Db 19946 AGAACAGCAGCAAGCGCGCAGATCCAGCTCCGCGTCACTGTGAGGTGAGCGCCC 19887
QY 808 ProProProAlaProProSerPro-----SerAlaProProProValValProLys 824
Db 19886 CGTCCACCTTCGGCGCGCAGCACCGCGAGAGCGCTCGGCGTCTGGGCGGTGAGCAGC 19827
QY 825 GluGluLysGluGluGluThrAlaAlaAlaProProValGluGluGluGluGln--- 843
Db 19826 CGTGTGCGAGCAGCGCGCGCAGGTGCGAGCAGCGCGCTCAAGCAGCGCGCGTCTGATGC 19767
QY 844 LysProProAlaAlaGluGluLeuAlaAlaValAspThrGlyLysAlaGluGluProValLys 863
Db 19766 CGGCGCAGCACCGCGGACCTCTCCGCTTCAGACGTGCGCAGCAGCACCGTCCACCG 19707
QY 864 SerGluCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAla 883
Db 19706 TCTCGCGCGCGAGTCTCTCAGTGACTGACAGAGTCTCGGCGCGCGCGCTCCAGCC 19647
QY 884 GluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGlyGlyGlySerGlyArgAla 903
Db 19646 CGC-----GACGCGACGTCAAGCAAGGTGCTCACCCCGT--- 19611
QY 904 ThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAla 923
Db 19610 -----GCGCGCAGCAGGTGCGCGCAGTGCAGTGCAGCCAGTCCCTCGCGCGG 19557
QY 924 AspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeu-----Leu 938
Db 19556 TCACGAGCAGCGCGCGCAGGTCCAGCTCGCGGCTCTCGGTGAGTCTCTCCGCTG 19497
QY 939 SerPro---ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerPro 957
Db 19496 CGCGCTCGCGCGCACAGCGCGCGCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 19437
QY 958 GlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProIleGln 977
Db 19436 GCTCCGCGG-----CGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19398
QY 978 ValThrLysValHisGluProProArgGluAspAla----- 989
Db 19397 CGTCCAGCGGCTCGGTCCCATCATGATCAAGCGCGCGCGCGCGCGCGGTGCTCGCTGC 19338
QY 990 AlaProThrLysProAlaProAla-----ProProProProGlnAsnLeuGlnPro 1007
Db 19337 GCAGCGCGGCAACAAGCG 19278
QY 1008 GluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAla 1027
Db 19277 GCAGCG 19218
QY 1028 ProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPro 1047
Db 19217 CCAGCAGCGCGCTTGCGAGCAGCGACAGCG-----CCTGCGCTCGTCCGCT 19173
QY 1048 CysTrpThrSerGlyLeuProPheProValProArgGluValLysAlaSerPro 1067
Db 19172 -----CGTGGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19122
QY 1068 HisAlaProAspProSerAlaPheSerTyrAlaProGly----- 1081
Db 19121 CGACCG 19062
QY 1082 -----HisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArg 1098
Db 19061 GATCGGCAATCGCTCGCGCTCAGGCT-----CGGCGCGCGCGCGCGCGCGCGCG 19014
QY 1099 ProProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeu 1118
Db 1118

19013 CTTCCGCTGCACCGACGACACCA-----CGAGCGCGCGCTCCAGGGGGGTG 18966
QY 1119 GluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSer 1138
Db 18965 CCGCGA-----GCTCCACAGGTGGAAGTCACTCATATATATATGTTGGCCACGAG 18915
QY 1139 GluHisAlaLysAlaProValGlyProValThrMetGlyLeu----- 1152
Db 18914 CGTGAACGGCGGTCCGCGAGTGTCTCGCGCGTCCGCGCGCAGATGTAGAGCGCGATGC 18855
QY 1153 ---ProLeuProMetAspProLysLysLeuAlaProPheSer----- 1165
Db 18854 TCGCCACGGCT-----GCCCGTGGAGTCTGCACTTGCAGTTCGACGCGAAG 18813
QY 1166 ---GlyValLys-----GlnGluGlnLeuSerPro 1174
Db 18812 CGGTGCTCTGTGTGGCTCTCGTCTGGAGCTCTACCCGAGCCCGAAGGTCTGCTCGCCC 18753
QY 1175 ArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerVal 1194
Db 18752 CCGTCCGCTGACGCGCACGT-----CCGACACGAGCAACCGGCACGAGCA 18708
QY 1195 LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSer 1214
Db 18707 CTTCTGTCG-----CCCTGGCTCTCGAGAGGCGCGCACCTCG 18666
QY 1215 -----ThrArgValProSerAspSerAlaIleThr---TyrArgGlySerIleThrHis 1231
Db 18665 TGTGAGCGCGGTCTCATCAAC-GCGGGTGCAGCCATACGCTCCGCTCGCTGCTCTTC 18607
QY 1232 GlyThrProAlaAspValLeuThrLysGlyThrIleThrArg-----Ile-Il 1247
Db 18606 TCGTCTCCC-----GGCAGCACCCACCGACCGAAGTAGTGTGCT 18565
QY 1247 eGlyGluAspSerProSerArgLeuAspArg---GlyArgGluAspSerLeuProLysG 1266
Db 18564 TGGCGGACAGCTCCACGAGCCCTTGAACCGCGGCGCT----- 18525
QY 1266 yHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerVa 1286
Db 18524 -----AGTGGAGCCCGCTACGAGCGCGCT----- 18498
QY 1286 lThrGlnCysSerLysGluAspGlyArgSerSerGlyPro-----ProHi 1302
Db 18497 -----CGTAAAGCCCGACAGGTCCACCTCTCCGCGCGCGCG 18460
QY 1302 sGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIl 1322
Db 18459 ACAGGCCACGTTCGACAGCTCTGAGCGCGAGGATGCGCGGGGCTCTGCTCGTCAACA 18400
QY 1322 eser-----SerAlaSerIleGluGlyLeuMetGlyArgAlaIlePr 1336
Db 18399 CTCCGCTCGGTGCTGCACCCAGCGCGCTCTCCGGGGCTCTCAGGCTGCTGTACA 18340
QY 1336 oProGlu-----ArgHisSerProHisHisLeuLysGlu----- 1347
Db 18339 GCCCACCTCGCGCGCGCGCGCTCGGGCTCCGATCATCATCATCGCAGCGCGCG 18280
QY 1348 -----GlnHisIleArgGlySer-----IleThrG 1357
Db 18279 CGCCCTCGGCCAGCACCGCGCTCGGCCAGCGGTGAGCTCCGACGAGCACCGGCTGCCA 18220
QY 1357 nGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGlu-----Al 1375
Db 18219 CCGCGCGCGCGCGCGCGCTCCAGCTCCAGCTCCCGTCCCGGAGCACCACTGCG 18160
QY 1375 aLysLeuLeuLys-----ArgGluGlyThrPro----- 1385
Db 18159 CAAACCTCATGATCCCGAAGCCACGATGCTCCGCCAGCGACGCGCGCTGTGAACA 18100
QY 1386 -Pro-ProProProSerArgAspLeuThrGluAlaTyrLysThr-----GlnAla 1402
Db 18099 GGTGCGCTCGCGTCCGCGAGCTCTGTCGCGCGCGAGCGGTGTGCTGCTGCT 18040

QY	1402	euGlyProLeuLysLeu-Lys-----ProAlaHisGluGlyLeu 1414	QY	1693	lnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyL 1713
Db	18039	TGAGCCCCGCGAGCTCAGCTCGCTACGGCGCTTCGAGATATCCAGCCAGTAACGCTGCC 17980	Db	17037	-----CCGCTCCGGAACCGCGCTCTGCTCCAGAGCGCCGCGC 16995
QY	1415	ValAlaThrVallysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArg 1434	QY	1713	euSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleA 1733
Db	17979	GCTGAACCGGTACGTCGCGAGCT-----CCAGCACACGAC---CGC 17941	Db	16994	CATCCCGCGCACCTGGCTCCCTT-----GGCCGGGGAACACGAACACCACTT 16947
QY	1435	HisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly 1454	QY	1733	spLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrProAlat 1753
Db	17940	CGTGGCCCGACACACCGCTGTCAGTCCACCT-----CGT 17905	Db	16946	GCCTCGCGACGCGCGCTACCCGCG-----ACACGCGCGGTGGCGCGACCTCGCA 16893
QY	1455	ThrProLeuLysTyrAspThrGlyAlaSerThrGlySerLysHisAspValArg 1474	QY	1753	hrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHis 1773
Db	17904	GCCCCGTCAGTCAGTCCCGAGCGCTCTGTACAGCTGCGACACCCAC---CTTCGT 17848	Db	16892	CAGG-----CCCGACGCGCTCTCCGCTCGACACGCTCGC 16854
QY	1475	SerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAla 1494	QY	1773	erSerSerProLeu-----SerProGlyGlyProThrHisLeuThrL 1787
Db	17847	CGCGTCCAGGCTCCCGACACCA-----CCCCCTCGCGCTCCCGCACGCGTGTCA 17794	Db	16853	CGCAAGCACCATCGCGGTGAGCGAAGTGTCTCGGTGAGCGCCCTGTCTCTCACCAC 16794
QY	1495	AspAlaArgAlaLeuGluArgAla---CysTyrGluGluSerLeuLysSerArgProGly 1513	QY	1787	ysProThrThrThrSerSerGluArgGluArgAspArgAspArgGluArgAspArgA 1807
Db	17793	CGCGCATCGCCAGCACCGGTGCGCGCTACCTCCAGACACACCGTCCGCTCTCCA 17734	Db	16793	GTCCGACCATGCACTCCCGCTCTCCCGCGCATCGCAACAGCGGCATCGCAACACCTCCGTCGC 16734
QY	1514	ThrAlaSerSerSerGlySerIleAla-Arg-----GlyAlaProValI 1529	QY	1807	spArgGluArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleT 1827
Db	17733	GCAGCTTCGACAGCGCTCGTCCAGGCGCACCGTCCCGCGAAGGTTCGCGCACCATCT 17674	Db	16733	CCCCACCGAGGCTCGTCTCGCCCGCACAGCAACAGCGGCATCGCAACACCTCCGTCGC 16674
QY	1529	eValProLeuGlyLysProArgGln-SerProLeuThrTyrGluAspHisGly--- 1547	QY	1827	tpArgProGlyThrGluGlnSerSerGlySerSerGlySerSerGlyGlyGlyGlyGly- 1846
Db	17673	CGCGTCCAGCGCTCCCGCGCACACCTCTCTGTCACTGTCAGTGAATACGGCAGCT 17614	Db	16673	C-AACGGTTCGCGCATTCCTCCACCGCTCGCGCGCGCTCCACCGCGCTCTTCCA 16615
QY	1548	-----AlaProPheAlaGlyHis----- 1553	QY	1847	-----SerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerPro---- 1862
Db	17613	CGCTCGCTTCGCGCTGAGCGACGACAGCTTCGCTCTAGCTCGCGGACGAGCGCATCCA 17554	Db	16614	GGATGACATGCGCGT-CGTCCGCTGATCCCGAACGACGACGCGCGCGCGCGCGCC 16556
QY	1554	-----LeuProArgGlySerProValThrMetArgGluProThrProArgLeuG 1570	QY	1863	-----IleSerProArgThrGlnAspAlaLeuGlnArgProSerValLeu-His 1879
Db	17553	TGTGCGCGCTGTGGACGCTAGTTCAGCTTCCTTCGCGCAACACACCTTCTGCCG 17494	Db	16555	CGCGCGTTTCGCGCGCGCGCGCTTCGACCAACGACGACGCGCGCTCCCTCCCTCCAC 16496
QY	1570	lnGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProA 1590	QY	1880	---AsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu 1898
Db	17493	TCAGTCCACCATAGCCGCTCCCGCTCCAGCTCCAGTCCAGTCCAGCACACCGT-CGAGCTC 17435	Db	16495	CAATGTGCGGCTCGGCTCGCTCGCGTGCAGCGTCCGCGCGCGCACCTCGTGTGCATC 16436
QY	1590	fgGluIleAlaLysSer----- 1595	QY	1899	ArgSerThrSerSerSerProValArgProAlaAlaThrPheProProAlaThrHis 1918
Db	17434	GACGTGTTACGCGCGCTATCGCAAGCGCTCCCGGTACGCGCGCATCGCTCTGCACC 17375	Db	16435	GACAGCACCATCTTGAGCACACCGCGCGCGCA-----GGCCCCGCG----- 16388
QY	1596	---ProHisSerThrValProGluHisHis-----ProHisProI 1608	QY	1919	CysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeu 1938
Db	17374	TCCGACACGCGCGCTCGACACGACATATCGCCCGCATCCCGCGCGCTCGCGCGCGC 17315	Db	16387	TGTCCA---GGTTCGACTTCGACGCGCAAGGTACAGCGCGCGCTCGGCTTCGCGCCAG 16331
QY	1608	leSerProTyrGluHisLeuLeuArgGlyVal-----SerGlyValAspLeuT 1624	QY	1939	ProLysGluAlaProArgValAlaArgProGluArg----- 1950
Db	17314	TGGTCCGCGCGCGCTACCCGCGCTCCCTCCGCAAGCGACGAGCGCTCCGACACCCACC 17255	Db	16330	GCCCCGAACACCTCCACGCGCTCCGCGCTCGATCGCGTTCGCGCGCGCGCGCGCTCC 16271
QY	1624	yrArgSerHisIleProLeuAlaPheAspPro-----T 1635	QY	1951	-----ProArgAlaAspThrGly-----HisAlaPheLeu 1960
Db	17254	GCGCGCGACACCTCGCGCTGGGTGT-GGCCCGACCGCGCGCGCGCTCGCGCGCGCG 17196	Db	16270	CATCGCGCTCCACCGGTGATGTCTCGGCGCAAGCGCGCGCGCGCGCGCGCGCGCTGT 16211
QY	1635	hrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisL 1655	QY	1961	AlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGlu 1980
Db	17195	CGCCACG 17136	Db	16210	GGATCACCGCTGCTGGCGAGCGCGTTCGCGCGCGCTCAACCTCGCTGGCTCGCGCGCT 16151
QY	1655	euAlaProAsnProThrTyrPro-----HisLeuTyrProTyrTyrLeuIleArgGlyT 1673	QY	1981	ProArgProLeuVal-----ProProValSerGlyHis----- 1991
Db	17135	CGGCTCCAGCGCGGTGCTCTCCCGCGCATCTCCCGCGACGACCGACGACGACGACCA 17076	Db	16150	GGTTCACCGTGAACCGGATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16091
QY	1673	yrProAspThrAlaAlaGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerG 1693	QY	1992	-----AlaThrIleAlaArgThrPro-----AlaLysAsnLeuAla 2003
Db	17075	GCCGCTCCAGCG 17038	Db	16090	ACAGCGGTTTCAGCACCGGATCCCGCACCCCTCGGACCGACGATGACGCGCGTCACTCGCG 16031
			QY	2003	----- 2003

Db 16030 CAGAGAACTCTTGTCAACGACCGTCCGCTGCAAGGGCTTTGAGCGGCTGAACCTCCACGA 15971
Qy 2004 -----ProHisHisAlaSerProAspProProAlaPro 2014
Db 15970 ATCCCGGGTGTGGTTCATCATCATCATCCACCGGTCAAGGCCAGATCGCACTCGCCCT 15911
Qy 2015 ProAla-----SerAlaSerAspProHisArgGluLysThrGlnSerLys 2029
Db 15910 GAGCAGGGCGGTGCACCGCCAGGTGCAGCGACACACAGCAGCAGCAGCGCGGTGTCCA 15851
Qy 2030 ProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSer 2049
Db 15850 CCG-----TGATCGCTGGGCGCTTCGAGCCCGAGCAGCAGTACG 15815
Qy 2050 ProGluGlyValGluProValSerProValSer---SerProSerLeuThrHisAspLys 2068
Db 15814 CCACAGCGCTGAGACGACGCTCGCCCGGTCCCGGTGCCCGGT----- 15770
Qy 2069 GlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgPro 2088
Db 15769 AGCGTCAAAGCGCGCAGGTGCTGCGTAAAGAGCAGCAGTGTCCGAGC-----CCA 15716
Qy 2089 LysGlnProGlyProValLysLeuGlyGluAlaAlaHisLeuProHisLeuArgPro 2108
Db 15715 TCGAGCCCGAGGTACACCCCGGTGGAGCTCTCGCTCAGCGCGCAGCGCGCAGCGCC 15656
Qy 2109 LeuProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGly 2128
Db 15655 GCTCCA---GTGCTCCACCGCGCTCTCCA-----GCACGAGCGGTGCTGGG 15611
Qy 2129 HisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThr 2148
Db 15610 GATCCATCGCTCGCGCTCTCGGGCG-----ATATCC 15578
Qy 2149 ArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGly 2168
Db 15577 CGAAGAAACCGCATCGA---ACAGGTGATGTCTCCGCGCAAGATCCACCTCGCGCAGT 15521
Qy 2169 AlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuProPro 2188
Db 15520 AGGTCTTGCCCA-----CCGCTCCG 15500
Qy 2189 AspHisGlyAlaProAlaArgGlySerProHisSerGluGlyLysArg----- 2205
Db 15499 GATCGGGTGTAAATCGAAGGTGCATCCACGCGCGCAGGAGGCTCGATCGCATCTC 15440
Qy 2206 SerProGluProLysThrSer----- 2213
Db 15439 TCCCTCTCTCCAGAGCGCAGTAGTCTTCCGGCGTAGCGAGCGCCCGCGCGCGC 15380
Qy 2214 -----ValLeuGlyGlyGlyGluAspGlyIleGluProVal 2225
Db 15379 ACCCATCGATAGTGGCGATGGCGCTCAACGAGATCCGATGGCTGGCTGCAGCC 15320
Qy 2226 SerProGluGlyMet-----ThrGluProGlyHisSer----- 2237
Db 15319 TCTGCAACTCTTCAATGCAAGGTGCAACGCTCAACGAGGTAGGAGAGCGCGCTGC 15260
Qy 2238 -----ArgSerAlaValTyrProLeuLeuTyrArg-----Asp 2248
Db 15259 TATTCATGGAGATCCTCGAGTTCGAATCATTTGCAAGGCTTCAAGGCGCAGCTGCTCG 15200
Qy 2249 GlyGluGlnThrGluProSerArgMetGlySer----- 2259
Db 15199 TCGCGAGATCTTCGAATCGGAATCGGAACCGGAACCTCGAGCGATGTTTCGGAGCGC 15140
Qy 2260 ---LysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSer 2278
Db 15139 CCGCGAGCTTCAAGACTTGGGAGGAGGCTGACGACGACGAGACTGACGAGGAGG 15080
Qy 2279 AsnSerAlaMetValLysSerLysLysGlnGluLeuAsnLysLysLeuAsnThrHisAsn 2298
Db 15079 CTGAGCATGAACCGCGGAATTTCTCTCGCATGAGTCCGCTGTACGAGCAGATTC 15020
Qy 2299 ArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAla 2318
Db 15019 CGCGCTCACCGTTTCGAACCTTGACAGAGGAC-----TTCCGATGTGGTG 14969
Qy 2319 IleThrGly-----ThrGlyLeuMetThrTyrArgSerGlnAlaVal 2332
Db 14968 GCAGATGGATGTCGAAAAGCAAAAGTGGCAGCAAGTTGCTTCGACGACGAGCGCAGC 14909
Qy 2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly 2352
Db 14908 CGATTCGGACTTCGACGGCCATG----- 14885
Qy 2353 LysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372
Db 14884 -----ACGGAATTCATTCAGGCTCCCTCAAGGCTGGCGGTAGGATC 14840
Qy 2373 AlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHis 2392
Db 14839 GCGTCA-----CGCGCTGAAGCCCGAGCACCAGCGGACCTCGGCCGCAAGG 14789
Qy 2393 ThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArg 2412
Db 14788 TCATGAGCGAGCTC-----AGCGCTCCGCTCCGCGCAGCGCGAGAGCGC 14741
Qy 2413 LysAlaLysSerProAlaProGlyLeuAlaSerGlyAsp-----ArgProProSerVal 2430
Db 14740 TCGCGAGCGCTCGAGGCTCTCTCTCGCGCAGAGACGCTTGCAGCGCCCGGGCGC 14681
Qy 2431 SerSerValHisSerGluGly-----AspCys 2439
Db 14680 ACCAGCGCGAAGCAGCGCGCGAGCTCGCGCTGGACTCCAGCCCGCGCTGCAGCTGC 14621
Qy 2440 AsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySer 2459
Db 14620 GCGAGATCCAGGTGCATCGGACCGAGCTCGCTTCAGAGCGTGAAGCGCGCTCAGC 14561
Qy 2460 ThrProPheProTyrAsn-ProLeu-----IleMe 2469
Db 14560 AGAGCGAGCGCTCGTTCGACCGATATCGGCAGCGCCTCGCGCTGTAGCGCAGT 14501
Qy 2469 thrGluGlnAlaGlyValMetAlaSerProProPro-----ProGlyLeuProAla-- 2486
Db 14500 TCGCGCTCGCCAGGTGCGCTGTCATCCCGCAGCGCTTGGCGCCACACCCCGCGC 14441
Qy 2487 -GlySerGlyProLeuAla 2492
Db 14440 AGGCTCGTCCGCGAGCC 14422
RESULT 56
AC114730/c 193894 bp DNA linear PRI 02-NOV-2002
LOCUS Homo sapiens BAC clone RP13-925E23 from 2, complete sequence.
DEFINITION AC114730
ACCESSION AC114730.5 GI:23463016
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 193894)
AUTHORS Tomlinson,C., Bielecki,L. and Doebber,A.
TITLE The sequence of Homo sapiens BAC clone RP13-925E23
JOURNAL Unpublished (2001)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 193894)
AUTHORS Tomlinson,C., Bielecki,L. and Doebber,A.
TITLE The sequence of Homo sapiens BAC clone RP13-925E23
JOURNAL Unpublished (2001)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 3 (bases 1 to 193894)
AUTHORS Waterston,R.H.

Mon Apr 19 08:15:02 2004

TITLE
JOURNAL
Submitted (11-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
4 (bases 1 to 193894)
Waterston,R.H.
Direct Submission
Submitted (07-MAY-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
5 (bases 1 to 193894)
Waterston,R.H.
Direct Submission
Submitted (02-OCT-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
6 (bases 1 to 193894)
Waterston,R.
Direct Submission
Submitted (02-NOV-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 2, 2002 this sequence version replaced gi:20486437.

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu

Summary Statistics

Center project name: H_FH0925E23

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-13 Human Female BAC Library was constructed using improved
cloning techniques developed by Kazutoyo Osoegawa. The library was
generated by Baohui Zhao in our laboratory. Construction was
funded by a grant from the National Human Genome Research Institute
(NHGRI, NIH) (H1R01RG01165-03). The library was generated according
to the new NHGRI/DOE 'Guidance on Human Subjects in Large-Scale DNA
Sequencing'.
Female blood was obtained via a double-blind selection protocol.
Female blood DNA was isolated from one randomly chosen donor (out
of 10 female donors) and partially digested with a combination of
EcoRI and EcoRII Methyase for library segments 1k2 or either MboI
or DpnII for library segments 3k4. Size selected DNA was cloned
into the pBAC3.6 vector between the EcoRI sites for library
segments 1k2 or the BamHI sites for library segments 3k4. The
ligation products were transformed into DH10B electrocompetent
cells (BRL Life Technologies). The library has been arrayed into
384-well microtiter dishes and also gridded onto 22x22cm nylon high
density filters for screening by probe hybridization.

NEIGHBORING SEQUENCE INFORMATION:
There is an unresolved tandem repeat from base 25865 to 27600.

There is an unresolved tandem repeat from base 39010 to 43450.
There is an unresolved tandem repeat from base 81306 to 83536.
Polymorphisms have been identified between AC133781 and this
sequence.

Data from AC133781 was used to finish this clone.
This sequence is the entire insert of the clone.

FEATURES	source	Location/Qualifiers
	1..193894	/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/chromosome="2"
		/map="2"
		/clone="RP13-925E23"
		/clone_lib="RPCI-13"
repeat_region	1..243	/rpt_family="Alu"
repeat_region	306..414	/rpt_family="L1"
repeat_region	543..673	/rpt_family="(TG)n"
repeat_region	1274..1526	/rpt_family="Alu"
repeat_region	1707..1727	/rpt_family="GC-rich"
repeat_region	1987..2074	/rpt_family="C-rich"
repeat_region	2856..3211	/rpt_family="L1"
repeat_region	3416..3453	/rpt_family="ERV1"
repeat_region	3454..3759	/rpt_family="Alu"
repeat_region	3760..4103	/rpt_family="ERV1"
repeat_region	4325..4773	/rpt_family="L1"
repeat_region	4778..4982	/rpt_family="ERV1"
repeat_region	4983..5280	/rpt_family="Alu"
repeat_region	5281..5394	/rpt_family="ERV1"
repeat_region	5395..5691	/rpt_family="Alu"
repeat_region	5692..5874	/rpt_family="ERV1"
repeat_region	5875..6007	/rpt_family="Alu"
repeat_region	6008..6279	/rpt_family="Alu"
repeat_region	6280..6398	/rpt_family="ERV1"
repeat_region	6399..6689	/rpt_family="Alu"
repeat_region	6690..6781	/rpt_family="ERV1"
repeat_region	6782..7072	/rpt_family="Alu"
repeat_region	7073..7343	/rpt_family="ERV1"
repeat_region	7408..7818	/rpt_family="ERV1"
repeat_region	7890..7986	/rpt_family="Alu"
repeat_region	7987..8027	/rpt_family="Alu"
repeat_region	8073..8220	/rpt_family="ERV1"


```
Db 42359 --CCCTCCCTCCCTCAGCCAGGACCTCGCAGGACACCTCCCTCCCTCCCTCAGCCAGG 42302
Qy 1111 SerAlaIysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSer 1130
Db 42301 GACCTCCGAGGACACCTCCCTC-----CCCTCAGCCAGGACCTCT 42257
Qy 1131 ValGlnLeu---HisValProTyrSerGluHisAlaIysAla-----ProValGlyPro 1147
Db 42256 GCAGGCTCCCTCCCTCCCTCCCTCAGCCAGGACCTCGCAGGACACCTCCCTCCCTCC 42197
Qy 1148 ValThrMetGlyLeuProLeuProMetAspProLysLeuAlaPro-----PheSer 1165
Db 42196 TCAACCCAGGAGCTCCGAGGACAC--CTCTCTCCCTCCCTCAGCCAGGACCTTCGCA 42138
Qy 1166 GlyValIysGlnGlnLeuSerProArgGlyGlnAlaGly----- 1179
Db 42137 GGCACCTCCCTCCCTCCCTCAACCCAGGACCTCGCAGGACACCTTCCTCCCTCCCTCC 42078
Qy 1180 ---ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThr 1198
Db 42077 AACCCAGGACCTCGCAGGACCT-----CCCTCCCTCCCTCAGCCAGGAC 42027
Qy 1199 AlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPro 1218
Db 42026 TCCGAGGACCTTCCTCCCTCCCTCA--ACCAGGACCTCGCAGGACCTCCCT 41970
Qy 1219 SerAspSerAlaIle---ThrTyrArgGlySerIleThrHisGlyThrProAlaAspVal 1237
Db 41969 CCCCCCTCAGCCAGGACCTCTGAGGCTCCCTCCCTCCCTCAGCCAGGACCTC 41910
Qy 1238 LeuTyrIysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArg 1257
Db 41909 CGCAGGACCTCCCTC----- 1277
Qy 1258 GlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysGlyHisVal 1277
Db 41870 CGCAGGCTCCCTCCCTCCCTC-----CTCAGCCAGG 41838
Qy 1278 LeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSer 1297
Db 41837 -----ACCTCCGAGGACCTCCCTCCCTCCCTCAGCC 41805
Qy 1298 SerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArg 1317
Db 41804 CAGGACCTCCGAGGACCTCTCCCTCCCTCCCTCAGCC-----CAGGACCT 41757
Qy 1318 ValGlyArgAlaIleSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProPro 1337
Db 41756 CTGAGGACCTTCCTCCCTCCCTCAACCCAGGACCTCTGAGGACCTTCCTCCCTCC 41697
Qy 1338 GluArgHisSer-ProHisHisLeuLysGluGlnHisIleArgGlySerIleThrG1 1357
Db 41696 CCC-----TCAGCCAGGACCTCCGAGG----- 41672
Qy 1357 nGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLe 1377
Db 41671 -----CAACCTCCCTCCCTCCCTCAGCCAGGAC---CTCCGAGGACCTCCCT 41622
Qy 1377 u-----LeuLysArgGluGlyThrProProProProProProSerAr 1391
Db 41621 CCCCCCTCAGCCAGGACCTCGCAGGACCTCCCTCCCTCCCTCAGCC-----AG 41568
Qy 1391 AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHi 1411
Db 41567 GACCTCCGAGG-----CAACCTCCCTCCCTCCCTCAG-- 41534
Qy 1411 sGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgG1 1431
Db 41533 ----- 41526
Qy 1431 uGluLeuArg-HisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySer1 1451
Db 41525 CCTCTGAGGACCTCCCTCCCTCCCTCCCTCCCTCAGCCAGGACCTCCGAGG----- 41473
```

```
Qy 1451 leThrGlnGlyThrProLeuLysTyrAspThrGlyAlaIleSerThrThrGlySerLysLysH 1471
Db 41472 --CCACCTCCCTCCCTCAGCCAGGACCTCGCAGGACCTCCCTCCCTCCCTCC 41419
Qy 1471 isAspValArgSerLeuIleGlySerProGlyA-gthPhPro-ProValHisProLeu 1490
Db 41418 -----TCAGCCAGGACCTCGCAGGACCTCCCTCCCTCCCTCAGCCAGG 41370
Qy 1491 AspValMet-----AlaAspAlaArgAlaLeuGluArgAlaCysTyr 1504
Db 41369 GACCTCCGAGGACCTCCCTCCCTCCCTCAGCCAGGACCTC-----TCAGG 41319
Qy 1505 GluGluSerLeuLysSerArgProGlyThrAlaIleSerSerGlyGlySerIleAlaArg 1524
Db 41318 CCACCTCCCTCCCTCAGCCAGGACCTCGCA----- 41283
Qy 1525 GlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGlu 1544
Db 41282 GGCCTCCCT-----CCCTCCCTCCCTC 41262
Qy 1545 AspHisGlyAlaProPheAlaGlyHisLeuProArgGlySer-----ProValThr 1561
Db 41261 ---AACCCAGGACCTCGCAGGACCTCCCTCCCTCAGCCAGGACCTTCGAG 41205
Qy 1562 MetArgGluProThrProArgLeuGlnGlySerLeuSerSerLysAlaSerGln 1581
Db 41204 GCCTCCCTCCCTCCCTCAG----- 41184
Qy 1582 AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHis-Ser----- 1598
Db 41183 -----CCAGGACCTCCGAGGACCTCCCTCCCTCCCTCCCTCAG 41145
Qy 1599 -----ThrValProGluHisProHisProIleSerProTyrGluHisLeuLeuArg 1616
Db 41144 CCAGGACCTCGCAGGACCTCCCTCCCTCCCTCCCTCAGCCCA----- 41102
Qy 1616 gGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSe 1636
Db 41101 -GGGACCTCCGCA-----GGCACCCTCCCTCCCTCCCTCAGCCCA----- 41063
Qy 1636 rIleProArgGlyIleProLeuAspAlaAlaIleTyrTyrLeuProArgHisLeuAl 1656
Db 41062 -----GGGACCTCCGAGGACCTCCCTCCCTCCCTCAG 41028
Qy 1656 aProAsnProThrTyrProHis-LeuTyrProProTyrLeuIleArgGlyTyrProAsp 1676
Db 41027 CCAGGACCTCGCAGGACCTCCCTCCCTCCCTC----- 40993
Qy 1676 hrAlaAlaLeuGluAsnArgGlnThrIleLeuAsnAspTyrIleThrSerGlnGlnMetH 1696
Db 40993 ----- 40993
Qy 1696 isHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArg-----GlyLeuS 1714
Db 40992 -----CAGCCAGGACCTCCGAGGACCTCCCTCCCTCC 40959
Qy 1714 erProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleAspL 1734
Db 40958 CTCCCTCCCTCAG-----CCGAGG-----GACC 40938
Qy 1734 euSerGlnValProHisLeuProValLeuVal-----ProProThrP 1748
Db 40937 TCCGAGGACCTCCCTCCCTCCCTCAACCCAGGACCTCCGAGGACCTCCCTCCCTCC 40878
Qy 1748 roGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnPro- 1767
Db 40877 CCCTCAACCCAGGACCTCT-----CGAGGACCTTCCTCCCTCCCTCAACCCAG 40827
Qy 1768 --PheSerSerArgHisSerSerSerProLeuSerProGly-----GlyP 1782
Db 40826 GGACCTCCGAGGACCTCCCTCCCTCCCTCCCTCAGCCAGGACCTCCGAGGACCTCC 40767
```


Db	39166	ACCTCCCTCCCTCCAGCCCA-----GGGACCTCCGACGACCACTCTCCCTCCC	39116	DB:	10	Gaps:	28
Qy	2479	roProProGlyLeuProAlaGlySerGlyProLeu-----AlaGlyProHisHia	2497	US-09-522-753-5 (1-2517) x MMU22016 (1-2036)			
Db	39115	CCCTGAGCCAGGACCTCGCAGGACCACTCTCCCTCTTAGCCAGGACCTCGCAGG	39056				
Qy	2497	la 2497					
Db	39055	CC 39054					
RESULT 57							
MMU22016							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
MEDLINE							
PUBMED							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
FEATURES							
source							
gene							
CDS							
polyA_site							
ORIGIN							
Alignment Scores:							
Pred. No.:							
Score:							
Percent Similarity:							
Best Local Similarity:							
Query Match:							

[illegible]

Db	281684	GCCACCGCGTGTGAGCAGGAGCCGCCCGCGGACACGTGCGTGTACGCGGGGCACACGTCACG	281622
Qy	538	-----ValGluAsnAspLysGlu	543
Db	281624	CGTCACCTCCACGGCAGCCGCCCGCTCGCGCACCGCGTGTGTCAGGGGCTGGAGACC	281565
Qy	544	AspLeuLeuLysGluLysThrAspAsp-Thr-----SerGlyGluAspAsnAspG	560
Db	281564	ACCGAGTCCGCGAGCGTGCCTGGCGAACACGTTGACGTGCAGCGGCACGCGTCTCCG	281505
Qy	560	wLysGluAlaVal-----AlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArg	578
Db	281504	CGCGCCCGCGCAGCAGCGCCGACCGCCAGCAGCGCGCTCGAGTTCGGATCGCGCG	281445
Qy	578	gLySGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrPr	598
Db	281444	CGCTCGCGCAGCAGTGCCTCGCGTCTCGGCGTCCGCGAGTATCTCCAGCGCGGACGCC	281385
Qy	598	oGln-----GlnSerAlaGluLeuAlaSerMetGluLeuAs	610
Db	281384	GCCGTCATCAGGTTGACCAACCGTTCGGAAGCGGAGCGGAGTGTCCGTCGAGACGG	281325
Qy	610	nGluSerSerArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHi	630
Db	281324	CACGGGACAGGATGG-----CGCGCAGGCTCCCCAA	281292
Qy	630	sGlyArgAsn-----TrpSerAlaIleAlaAr	639
Db	281291	CGGCGCGGTTCACGCGACGTTACGCGAGGATTCGGTGCCTACGCGGGGTGTGCCG	281232
Qy	639	gMetValGlySerLysThrValSer-----GlnCysLysAsnPh	652
Db	281231	ACGGTTCGCGCGCGGCCACTTCCTCCCGGAGCCCGGCTGTGATAATGCGAGGAACA	281172
Qy	652	eTyrPheAsnTyrlsLysArgGlnAsnLeuAspGluIleLeuGlnGln-----	668
Db	281171	T-----GACAGACTTCGCGCGCCCTCCCGCGTC	281142
Qy	669	HisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLys-LysAlaP	687
Db	281141	TGGCCACCACCTTCGCGCCCTGTGTCGCGTCAC--CGCGTCCCGCGCGTCCGCGC	281085
Qy	687	roAlaAlaAspGluGluAlaAlaPheProProValValGluAspGluGluMetGluA	707
Db	281084	CGGCAACGCG-----CAGGCACGGCACGACCGCGAGCCCCCGAGGACTTCGTGGGCC	281031
Qy	707	LaSerGlyValSerGlyAsnGluGluMetValGluGlu-----	720
Db	281030	TGCGCGGCGTC-----GACCCGACGATCATCGAGGAGATGCGTACGCGCGC	280980
Qy	720	-----	720
Db	280979	ACAACTTCGTCGGGAGCGGTTCGACGCGTACCGGCAGCCCCGTGTGCATCTCTCACCCGCGC	280920
Qy	721	AlaGluAlaLeuHis-----	725
Db	280919	CGGCGCGCGAAGCGCTCCACTTCGCCCCAGACCCGCGTCTGTGGCGGAGGCTACTCCTCA	280860
Qy	726	AlaSerGlyAsnGluValPro---ArgGlyGluCysSerGlyProAla	740
Db	280859	AGGTGTACACTGTCTACCGGCCCGCGCGCGTGCACCACTTCGTGCGTGGCGCGAAG	280800
Qy	741	ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaLys	760
Db	280759	ACCTCGACGACGAGACATGAAGCGGAGTTCTACCCGCGACGTCGACAGACCCGCTCT	280740
Qy	761	AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAsp-----	776
Db	280739	TCGCCGACGGCTACATCCCGGAGAGTCCGGCCACACGCGCGGCTCCACCGTGGACCTCA	280680
Qy	777	GlyProProGlyProPro-----ThrPro-----	785
Db	280679	CCCTCGTACGCTCCCGGCCAAACCAACCGCGCGTACCAACCCCGGCGAGCCCTGGTGC	280620


```
Qy 786 -----ProArgArgThrSerArgAlaProIleGluPro----- 796
Db 280619 CCTGTCGCCCGCCGAAGGGGAGCGTTTCCTCCGACAACTCGTCACATGGCGCACCGGAT 280560
Qy 797 -----ThrProAlaSerGluAlaThrGlyAla 805
Db 280559 ACGACTGTCGACACCGCGTCGACACCCCTCGACCCCGCATCGAGGGCGTCCAGCGTG 280500
Qy 806 ProThr----- 807
Db 280499 CCAACCGGCTGCTGCTCAAGACACCTGGAGGAGGGCGGCTGGTGAACCTCCCGAGG 280440
Qy 808 -----ProProProAlaProProSerProSerAlaProProProValValProLys 824
Db 280439 AATGGTGGCACTTCACCTACAGGCCCGAGCGCTACCGGACACCTACTTCGACTTCGCG 280380
Qy 825 GluGluLysGluGluGluThrAlaAla----- 834
Db 280379 TTTCGCGGAAGTCTCTCCGCGCGCGCTGACCGGCTGCGGAGCGCTCCACTCCGCGC 280320
Qy 835 -----ProPro-ValGluGluGluGluGluGluLysProProAlaAla----- 848
Db 280319 ATCGGATACAGTCCGCGCGTGTCCGAAATTGAAAACTCCGTTCCCACTCCGCGCTCGG 280260
Qy 848 ----- 848
Db 280259 CTCCCACTGTTGAGCTGCGGATCGCCTACGGAGAGGGGCTCTCCGCTGGCCCCGAC 280200
Qy 848 ----- 848
Db 280199 CTGTCCCGGGTGGCGGACCGTGGCTACCGCAACCCGCTGCGGTCGCGGTGGCGCTCCA 280140
Qy 848 ----- 848
Db 280139 GCCGTGTACACAGCAGGGCACCGGCTGCTGCTCATCACCCGACCGTGACCCCGCGC 280080
Qy 848 ----- 848
Db 280079 GCGGGGGGGTGCCCTGCCCGCGGCTACGTCGACGACCGGGAGGACTGGCGGCGAGC 280020
Qy 849 -----GluGluLeuAlaValAspThrGlyLysAlaGluGluProVa 862
Db 280019 CGTGTCCGCGAGCTCAAGGAGGAGACCGGCTCGACGC-GGCGAGCGCGACGTCGCGC 279961
Qy 862 LysSerGluCysThrGluGluAlaGluGluProAlaLysGlyLysAspAlaGluAl 882
Db 279960 TCGCCGAGCG---CATGAGCTGCGCCGACGGCCACCTGCTGCTCTTCGCGCTGCTCCCG 279904
Qy 882 aAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGluGlyGlySerGly-- 901
Db 279903 AGCGCGCGCGGAGGAGTGCCTGCTCCGCGCGCACGAGACGCG-GGGGCTGGCAC 279845
Qy 902 -----ArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
Db 279844 CTCTCGCGAGTCGGAAGAGCTCGCTCTCCGCTCCACACGAGCGCGTACGCGCTTC 279785
Qy 916 pSerSerAlaThrCysSerAlaAspGluAlaValAspGluAlaGluGly-----GlyAs 933
Db 279784 TTCAGGCGCGTACATCTGAGCGCGAGCGTCCCGGAGCCACGAGCGGACCGCGGACG 279725
Qy 933 pLysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAl 953
Db 279724 GACGGCTCGCTCAGACCGTCTCTCCCTCCCGCT-----CCACGACG 279683
Qy 953 aAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaI 973
Db 279682 ACCCGCGCGCCT-----CCACGCGGACACGTACCGTTCGA-----TCTCCGGT 279638
Qy 973 eProProIleGlnValThrLysValHisGluProProArgGluAspAlaProThrLy 993
Db 279637 TCACACCGCGCTGCGCGCTCGGACGACCACTCCACCGCGCGTCCGCGCGCGCGCGC 279578
```

```
Qy 993 s----ProAlaProProAlaPro-----ProProProGlnAsnLeuGlnProGluSe 1009
Db 279577 GGGCGCCACACCTCGAGCTCGGCGCGCGCTCGCCGCCA-----GGACCGGAATC 279527
Qy 1009 rAspAlaPro-----GlnGlnProGlySerSerProArgGlyLysSerArgSerProAl 1027
Db 279526 ACCGCTCCCGCGCGCGAAACACCGGAATCGCGCCGAAGCGCGTCCAGCACCTGCG 279467
Qy 1027 aPro-----ProAlaAspLysGluAlaPheAlaAlaGluAl 1039
Db 279466 GCCGCTCTTCGTACGCGCGCTCGCTCGCCACGCTGTCACAGCGCGCCCGGGCAACCGC 279407
Qy 1039 aGlnLysLeuProGlyAspProProCysTTrpThrSerGlyLeuProPheProValProPr 1059
Db 279406 ACCGCAC---GCCGCTCGCGCGCGCTCGAGCACCGGGGCGCACGACAGTCCGCC 279350
Qy 1059 oArgGluValLysLysAspProHisAlaProAspProSerAlaPheSerTyrAlaPr 1079
Db 279349 AGCAAAACCGCTCTCGAGTCCCGCAGCATCCGCTCTCCGGCGTCCGACCCACAGG 279290
Qy 1079 oProGlyHisProLeuGlyLeuHisAspThrAlaArgProValLeuPro-ArgP 1099
Db 279289 GGACCGCAC-----AGGAGACACCCGTACGCGGGCGAGTGGCGCAGCGTC 279242
Qy 1099 roProThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuG 1119
Db 279241 ACGAAGTACGCGACGCGCGCGCTC-----GACGAGCGCACGCGCGCTGCTCC 279188
Qy 1119 luArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerG 1139
Db 279187 AGCAC----- 279183
Qy 1139 luHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProL 1159
Db 279182 -----CTCGTCCCCGAATCCACCGCTCCCGTGCGC----- 279150
Qy 1159 ysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaG 1179
Db 279149 -----GGCGCGCGCGCGGTGGTGGGAACAGCGCAGGTACCGCGCAGCTGC 279098
Qy 1179 lyProProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGly--T 1198
Db 279097 AACACCGCGAG-----ATACAGCTCCCGGGAAGAA 279068
Qy 1198 hrAlaLeuGlySerValProGlySerIleThrLysGlyIleProSerThrArgValp 1218
Db 279067 CCCCCCTCGAGCGG---CCGCGATCCGCGCCGAGACGCGCACCCGCGCAGCGGCC 279010
Qy 1218 roSerAspSerAlaIleThrTyArgGlySerIleThrHisGlyThrProAlaAspValL 1238
Db 279009 CCAGCACCGTGGCAGCGCGCGCAACCCC-----GGCACCCCGTGGCGCTT- 278960
Qy 1238 euTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgG 1258
Db 278960 ----- 278960
Qy 1258 lyArgGluAspSerLeuPro-----LysGlyHisValIleTyrGluGlyL 1273
Db 278959 --GCCCGGACAGGTGCTCCCGTAGCGCTGCAGCGCGGCCACCCCGAAGCGGAGGAG 278902
Qy 1273 ysLysGlyHisValLeuSerTyrGluGlyMetSerValThrGlnCysSerLysGluA 1293
Db 278901 CGAAGCGC-----CGCTCGT 278887
Qy 1293 spGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspM 1313
Db 278886 CGGGGGCCAGTGGCCGAGCCCTCTGATCCGCGCCCGCCCAT-GCACGCGCGTAGACG 278828
Qy 1313 etMetGluGlyArgValGlyAlaIleSerSerAlaSerIleGluGlyLeuMetGlyA 1333
Db 278827 TTGTGCGCTCCCTGTG-----GTACCGCGCGCGCTCCAGCGCATCGCGGCC 278777
Qy 1333 rgAlaIleProProGluArgHisSerProHisLysLeuLysGluGlnHisIleArgG 1353
```

```
Db 278776 GA-----CGGGGACGCTCGACTCCCC-----GAAGGCGCGAACGAG 278738
QY 1353 lySerIleThrGlnGlyIleProArgSer-----TyrValG 1365
Db 278737 GTGGGTCGTCCAGGTGCTGCCAGAACCGGTTGAACCCCTGCCGAGCGCTCTCTCGTAG 278678
QY 1365 luAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGlyGlyThrPro- 1384
Db 278677 AGCCGCCCCACCACTCGCGCGCACCGCGGTGCTGAAGTCCGGGAAGACCACTCCCCC 278618
QY 1385 -----ProProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuG 1403
Db 278617 GGACGCCCCACCCCGCGAGCACCGT-----278592
QY 1403 lyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyA 1423
Db 278591 --CCCGCCCGCATCCCGCACGAACAC-----GTCCCGGCGCCACCG 278552
QY 1423 rgSerIleHisGluIle-----ProArgGluGluLeu-----ArgHisThrP 1437
Db 278551 CCCTGCTACAGAGTTGCGGGCGCGCGGCGCGAACCGCGGCTCCACGGCCGACACACGAG 278492
QY 1437 roGluLeuProLeuAlaProArgProLeuLysGlu---GlySerIleThrGlnGlyThrP 1456
Db 278491 CGGAC---CCCTCCCTCGCGAGCTCTCGCGCAGTACGGGCGAGCTTCGGGAA-----C 278441
QY 1456 roLeuLysTyrAsp-----ThrGlyAlaSerThrThrGlySerLysLysHisA 1472
Db 278440 CGCTCTCTCGTACGGGTGAACACTCGCGGCGAGTGCCTGCCGATGCCAGATGCACC 278381
QY 1472 spValArgSerLeuIleGlySerPro---GlyArgThrPheProProValHisPro----- 1489
Db 278380 CGCTCCAGGGCAGATCCC-GCTCTGTGTGCCGCGCACACCGCGCTCGAGTCTCTGCTC 278322
QY 1490 -----LeuAspValMetAlaAspAlaArgA 1498
Db 278321 GTCCGCCGAAATCCACAGCGGTGATGATACCGAGCGCCACCGCGGCGAGCGCGGG 278262
QY 1498 laLeuGluArgAlaCysTyrGluGluSerLeuLysSerArg-----ProG 1513
Db 278261 CCCCCGGTACGAGCGCCAGGCGCAGCAGCAGCGCGGGGTACCCACCGTCACCCA 278202
QY 1513 lyThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProVal-IleValProGlu 1532
Db 278201 GCACGCGAGCGGACCGCGTCCATCCGAGCAGCGTCCGCGCGCGCGGTGTCGCCCGA 278142
QY 1533 LeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGly 1552
Db 278141 CCGCGGCC-CTTCCTCACCTCCCGCAACG-----CCACCGTGCCTGCC 278098
QY 1553 HisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGly 1572
Db 278097 ACGTGTCTGTTGGGAAGACCAAGTGGCTCCCGCGTCCGCCACCAACCAAGTGCACCGCA 278038
QY 1573 SerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIle 1592
Db 278037 TCGTGATGACACCGGACGAGCGGTCTCTCGCGCGCGCGGAGGCGAGCGGACCGG 277978
QY 1593 AlaLysSer-ProHisSerThrValProGluHisHisProHis-----ProIleSerPr 1610
Db 277977 TGTTCACAGCGGTACGTGCTCCGCGCAGCGGGGCGCGCGCGCGCGCGCGAGCC 277918
QY 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIle--Pro 1629
Db 277917 CGAAGAA-----CCGGGCGTC-----CCGCGCACCTCCGCCC 277885
QY 1630 LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaIatyr 1649
Db 277884 GCTGCGTCCAGCGGGCGGAGCCCGCCCGGCTCCACCCACCGGGGCGGCGAGTCAAC 277825
QY 1650 TyrLeu-----ProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrPro 1666
```

```
Qy 1972 -----AlaSerSerProSerLysGlySerGluProArgProLeuValProValSer 1989
Db 276858 GGACGAACACCGCTCTGTAGTGCAGAGACCAACGA---ACCGG-CCCGCGTCACT 276803
Qy 1990 GlyHisAlaThrIleAlaArgThrProAlaLys-----AanLeuAlaPro----- 2004
Db 276802 GGGCAGAGCTCGCGCGCAGTGCCTCCCTCCGCGCGCAACTCGCGCGCTCGCGGTAC 276743
Qy 2005 -----HisAla 2007
Db 276742 GCGCGCGGACCGGTCAGCGGTACCTCCCAACATCCCGCAGCGCGTGCCTCGCGTGC 276683
Qy 2008 SerProAspProPro-----AlaProProLa-----Ser 2017
Db 276682 TCGCCACCGCGCGGTGGCGCGCTCTGGACCTCTCGCGCGCGACTTCGGCGCCCGCA 276623
Qy 2018 AlaSerAspProHisArgGluThrGln-SerLysProPheSerIleGlnGluLeuG1 2037
Db 276622 GCGTCTCGACCGCTTCCAGCAGGTGCAACCGTCTCTGTTCCACCGTCGACGGCTACC 276563
Qy 2037 uLeuArgSerLeuGlyTyHis-----GlySerSe 2047
Db 276562 GCTACGGGGGCAAGAAACACAGCCGCGGACACCGTCCGCGAACTCCGCGGGAAGTGC 276503
Qy 2047 rTySerProGluGly---ValGluProValSerProValSerSerProSerLeuThrHi 2066
Db 276502 CCACCTCCGCGCGGTATCCACATCCCGTCTCGGCACCGCGCGCGCGCGC 276443
Qy 2066 sAspLysGlyLeuProTyHisLeuGluGluLeuAspLysSerHisLeuGluGlyLe 2086
Db 276442 TGGACTGGGAGACCT---CACCGCGCGGACCGCGGTGTACGAGCAGGTGCCT 276386
Qy 2086 uArgProLysGlnProGlyProValLysLeuGlyGluAlaahisLeuProHisLe 2106
Db 276385 TCGACCAACCCCTGTGGTGTCTACTCTCCGGCACACCGCGCTCCCAAGGCCAT-- 276328
Qy 2106 uArgProLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAla---ProG1 2125
Db 276327 -CGTCCAGTCCCGGCGGATCTCTCGAGCACCT---CAAGCAGTCCGCGCTGC 276275
Qy 2125 yValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrG1 2145
Db 276274 ACTCGACCTCGGCGCGGACCGCTTCTGTGTACACCTCGACCGGCTGGATGATGT 276215
Qy 2145 nAspTyThr-----ArgHisHis-----ProGlnGlnLeuSerAl 2157
Db 276214 GGAATCTCTCGTCTCGGCTGTCTACCGGCACACGATGCTCTTACGACG-GCAGC 276156
Qy 2157 aProLeu---ProAlaProLeuTySerPheProGlyAlaSerCysProValLeuAspLe 2176
Db 276155 CCGGCTTCCCGCACCGACCGCCAGTGGCGCATCGGAGACGACCGCGCCACCTT 276096
Qy 2176 uArgArgProProSerAspLeuTyLeuProProAsp-HisGlyAlaPro---AlaA 2195
Db 276095 TTCGGACCT-----CGCGCGCTACGTCATGCTGCGCTCGCAAGGCC 276054
Qy 2195 xGlySerProHisSerGluGlyGly-----LysArgS 2206
Db 276053 GGCGTCCACCGCGCGCGACCTGAGCTCTCCCGATCCAGTGCAGTGCACACCGCGC 275994
Qy 2206 erProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProValS 2226
Db 275993 TCCCGCTCCCGCGCGAGGCTCCGCTGGCTGACGACGATTCGCGCGCGGGTGC 275934
Qy 2226 erProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTy-ProLeuLeuT 2246
Db 275933 GACCTGTGGATCGCTCCGTCAGCGCGCGGACCGACGT-----GTGCTCC 275889
Qy 2246 yArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGly-----A 2264
Db 275888 TGCTTCGCGGCGCGCTCCACCTCCCGTCCACATCGCGGAAGTCCAGGCCCCCGCGC 275829
```

```
2264 snThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValL 2284
275828 CTGGCACCACTCCA-----GTCTGG 275805
2284 ySerLysLysGlnGluLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluT 2304
275804 GACCCGAGCGCAACCCCTTGACCGAGGTCGGCGAGCTGGTGGTGCACCAACCCGATG 275745
2304 yAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyGlyL 2324
275744 CCCTCGATCGGATCCGCTTCTGGAACG-ACCCCGACGGCAGCGCTACACGACAGCTA 275686
2324 euMetThrTyArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluA 2344
275685 CTTTCGACA-----CCTACCCCGCGCTGTGGCGGACGCGGACTGGA----- 275645
2344 laIleIleArgLysAlaLeuMetGlyLysTyArgPheGlnTrpGluLysSerProLeuS 2364
275644 -----TCGTCATCCAGCGCGCTCCGACTCCACCTCAACCG 275584
2364 erAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleT 2384
275625 CTCCG----- 275584
2384 hrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaL 2404
275583 CCAAGGCGTCCGATGGCTCCGCGGACATCTACGAGCGCTCGAGCGGCTCCCGGAGA- 275525
2404 yValSerGlyArgProSerSerArgLysAlaLysSerPro-----AlaProGlyLeuA 2422
275524 -----TCAGGGAATCCCTGCTCATCGGCATCGAACACCGCGCGCTACTCGATGCC 275470
2422 laSerGly-----AspArgProProSerVal-----SerSerValHisSerG 2436
275469 CTTCTTCGTGCACCTGCACCCCGGAGCACCTCGACGACGCGCTCTCGGACCGCATCA 275410
2436 luGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerS 2456
275409 GCGGACCATCCGCGCAACCTCTCCCGCGCCACGCTCCCGACGAGTCTATCGAGGTCCC 275350
2456 erAlaGlySerThrProPheProTyRhanProLeuIleMetArgLeuGlnAlaGlyValM 2476
275349 CGGCATCCCGCACACCTCCACCGCAAC-----GCATCGAGGT 275311
2476 etAlaSer-ProProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHis 2495
275310 CCGGTCAACGGCTCTCTCAGGGCACCCGCTGGACAAGCGGTCAACCCCGCTCAT 275251
2496 HisAlaTrpAspGluGluProLysProLeuLeu 2506
275250 CGA-----CAACCTCGACCTGCTCCACTTCTA 275224

RESULT 60
AC106029 180591 bp DNA linear PRI 06-AUG-2002
LOCUS Homo sapiens chromosome 15, clone CTD-2538111, complete sequence.
DEFINITION AC106029
ACCESSION AC106029.8 GI:22094423
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Mammalia; Eutharia; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 180591)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
TITLE Anderson,S., Barna,N., Bastion,V., Boguslavsky,L., Boukhalter,B.,
JOURNAL Brown,A., Camarata,J., Campopiano,A., Chang,J., Chararo,B.,
REFERENCE Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
2 (bases 1 to 180591)
AUTHORS Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge.S., Faro,S.,
```



```

repeat_region complement(10259..10512)
repeat_region /rpt family="MER46C"
repeat_region complement(10635..10802)
repeat_region /rpt family="MIR"
repeat_region complement(11162..11476)
repeat_region /rpt family="AluSq"
repeat_region 11477..11963
repeat_region /rpt family="Trigger3 (Golem)"
repeat_region complement(11996..12161)
repeat_region /rpt family="MIR"
repeat_region 12270..12563
repeat_region /rpt family="AluY"
repeat_region 12959..12993
repeat_region /rpt family="GC_rich"
repeat_region 13464..13497
repeat_region /rpt family="AluSx"
repeat_region 14055..14171
repeat_region /rpt family="(TG)n"
repeat_region /rpt family="L2"
repeat_region 14364..14650
repeat_region /rpt family="AluJo"
repeat_region complement(14920..15228)
repeat_region /rpt family="AluSx"
repeat_region complement(15232..15518)
repeat_region /rpt family="AluSx"
repeat_region 15672..15743
repeat_region /rpt family="(TA)n"
repeat_region complement(15744..16059)
repeat_region /rpt family="AluSx"
repeat_region complement(16069..16195)
repeat_region /rpt family="FLAN_C"
repeat_region 16196..16224
repeat_region /rpt family="AT rich"
repeat_region complement(16411..16962)
repeat_region /rpt family="L1MB7"
repeat_region complement(16963..17126)

Alignment Scores:
Pred. No.: 1,6e-05 Length: 180591
Score: 715.50 Matches: 566
Percent Similarity: 32.70% Conservative: 307
Best Local Similarity: 21.20% Mismatches: 1112
Query Match: 5.41% Indels: 694
DB: 9 Gaps: 115

US-09-522-753-5 (1-2517) x AC106029 (1-180591)

Qy 14 AlaThrGluProArgTyrProHisSerLeuSerTyr-ProValGlnIleAlaArgTh 33
Db 68885 GCCAGCTGGCCAGACAGCAGCCAGCCACTCAGCCAGCCAAAGCCAGCCAGCCAGCCAGCC 68944
Qy 33 rHisThrAspValGlyLeuLeuGluTyrGlnHisSerArgAspTyrAlaSerHisLe 53
Db 68945 ACCAGT-----CAGCCAGCCACTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 68989
Qy 53 uSerProGlySer---lleGlnPro---GlnArgArgArgProSerLeuLeuSerGl 71
Db 68990 CCAGCCAGCCAGCCAGCCCTTCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69038
Qy 71 uPheGlnProGlyAsnGluArgSerGlnGluLeuHis---LeuArgProGluSerHis 90
Db 69039 ----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69092
Qy 90 rTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProArgLe 110
Db 69093 -----CCGCCCGGC-----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69121
Qy 110 uGluLeuLeuPro---AspProLeuLeuArgProSer-----ProLeuLeuAl 125
Db 69122 GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69181
Qy 125 aThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys-- 144
Db 69182 GGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69241

```

```

Qy 145 -LeuGluProValSerProProSerProProHisThrAspPro----- 158
Db 69242 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69301
Qy 159 -----GluLeuGluLeuValProProArgLeuSerLysGluLeuLeuGlnAsnMe 176
Db 69302 CCAAGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69341
Qy 176 tAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLeuLys 196
Db 69342 -----ACACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69394
Qy 196 sGlnGlnGlnLeuGluGluAlaAlaLysProProGluProGluLysProValSerPr 216
Db 69395 CACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69454
Qy 216 oProProIleGluSerLysHisArgSerLeuValGlnIleLeuTyrAspGluAsnArgLy 236
Db 69455 GCCAAGC---CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69493
Qy 236 sLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLe 256
Db 69494 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69539
Qy 256 uTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMe 276
Db 69540 ----AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69551
Qy 276 tArgLysLysLeuLeuLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysGl 296
Db 69552 -----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69582
Qy 296 nLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuLysLysValGluArgIl 316
Db 69583 A-----GCCCGGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69591
Qy 316 eGluAsnAsnProArgArgAlaLysGlySerLysValArgGluTyrTyrGluLysGl 336
Db 69592 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69628
Qy 336 nPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGl 356
Db 69629 ----CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 69684
Qy 356 nArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIl 376
Db 69685 CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69727
Qy 376 eAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIlePr 396
Db 69728 -----CTAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69780
Qy 396 oProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMe 416
Db 69781 AAGCCTGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69819
Qy 416 tAlaAspProMetLysValTyrLys-----AspArgGlnValMetAsnMetTr 432
Db 69820 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69879
Qy 432 pSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGl 452
Db 69880 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69938
Qy 452 yLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLe 472
Db 69939 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69951
Qy 472 uThrLysLysAsnGluAsnTyrLysSerLeuValArgSerTyrArgArgArgGlyLys 492
Db 69952 -----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 69959

```


QY	1843	yglyglySerSerSerArgProAlaSerHis-----AlaHisGlnHisE	1861
Db	73668	AAGCCAGCCAGCCA-AGCCAGCCAGCCAAAGCCAGCCAGCCAGCCAGCCAA	73726
QY	1861	rProIleSerProArgThrGlnAspAlaLeuGlnArgProSer-----ValLeuHi	1879
Db	73727	GCCAGCCAAACCCAGAGAGCAGCCAAATCCAGCTAGCAGCAGCAAGCAAGCCAGCCAGTCA	73786
QY	1879	sAsnThrGlyMetLysGlyLeuThrAlaValGluProSerLysProThrValLeuAr	1899
Db	73787	TGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	73846
QY	1899	gSerThrSerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCy	1919
Db	73847	AGCCAGCCAAATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	73895
QY	1919	sProLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeuLeuPr	1939
Db	73896	CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	73948
QY	1939	oLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPh	1959
Db	73949	CAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	74008
QY	1959	eLeuAlaLysProProAlaArg-----SerGlyLeuGluProAlaSerSerProSerLy	1977
Db	74009	CCAGTCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74067
QY	1977	sGlySerGluPro-----	1981
Db	74068	GGCAAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74127
QY	1982	ArgProLeuValPro-ProVal-----SerGlyHisAlaThrIleAlaArgThrP	1998
Db	74128	AGTCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	74187
QY	1998	roAlaLysAsnLeuAlaProHisAlaSerProAspProProAlaProProAlaSer-	2017
Db	74188	CAGCCAAAGTCA--GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74244
QY	2018	AlaSerAspProHisArgGlyLysThrGlnSerLysProPheSerIleGlnGluLeuG	2037
Db	74245	AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74282
QY	2037	luLeuArgSerLeuGlyTyHisGlySer-----	2048
Db	74283	CACTCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74322
QY	2048	yrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspL	2068
Db	74323	AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74369
QY	2068	ysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyGluLeuArgP	2088
Db	74370	-----CAGC	74373
QY	2088	roLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgP	2108
Db	74374	CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74433
QY	2108	roLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysG	2128
Db	74434	AAAGCCCGCCAAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74492
QY	2128	lyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrt	2148
Db	74493	-----GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74529
QY	2148	hrArgHisIleProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrsPheProG	2168
Db	74530	AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	74576
QY	2168	lyAlaSerCysProValLeuAspLeuArgProProSerAspLeuTyrsLeuProProp	2188

RESULT 61
AF124821

LOCUS	AF124821	1599 bp	mRNA	linear	ROD 27-MAY-1999			
DEFINITION	Rattus norvegicus nuclear receptor co-repressor (NCOR) mRNA, partial cds.							
ACCESSION	AF124821							
VERSION	AF124821.1	GI:4894625						
KEYWORDS								
SOURCE	Rattus norvegicus (Norway rat)							
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.							
REFERENCE	1 (bases 1 to 1599)							
AUTHORS	Boutell,J.M., Thomas,P., Neal,J.W., Weston,V.J., Duce,J.A., Harper,P.S. and Jones,A.L.							
TITLE	Aberrant interaction of transcriptional repressor proteins with the Huntington's disease gene product, huntingtin							
JOURNAL	Unpublished							
REFERENCE	2 (bases 1 to 1599)							
AUTHORS	Boutell,J.M., Thomas,P., Neal,J.W., Weston,V.J., Duce,J.A., Harper,P.S. and Jones,A.L.							
TITLE	Direct Submission							
JOURNAL	Submitted (26-JAN-1999) Huntingtons Disease Group, Medical Genetics, U.W.C.M, Heath Park, Cardiff CF4 4XN, United Kingdom							
FEATURES	Location/Qualifiers							
source	1..1599							
	/organism="Rattus norvegicus"							
	/mol_type="mRNA"							
	/db_xref="taxon:10116"							
	/tissue_type="brain"							
	/clone_lib="HybriZAP[TM] Two-Hybrid activator domain vector (J.M. Boutell, et. al., Hum. Mol. Genet. (1998) 7, 371-378)"							
	/dev_stage="adult"							
gene	<1..>1599							
	/gene="NCOR"							
CDS	<1..>1599							
	/gene="NCOR"							
	/codon_start=1							
	/product="nuclear receptor co-repressor"							
	/protein_id="AAD12566.1"							
	/db_xref="GI:4894626"							
	/translation="KKGPPPKSRYYEELTRGKTTITANFDIVITRIQIASDKAR ERQSGSDSSSSSHRYEAPDAIEVISPAFPQEPKPTQYPMVKANQANES PQYQEPKPLTHRSQQSPQPPPLPSSQAEQGQVPRTHRLITLADHICQITQD FARNQVSPSTFTQTSFSALESTPVRTKPSRYPESQTVLHPRGPRVSPNL VDKNSGRQPKSPERSHIDPEYEPISPPQPAVHEKQDSMLLSQMGDPAPQRSD RSPGSLYLFPYFTKLESTSPMYKKQEIFRKLNSGGGDSMAAAQGTETFLNLPATVTSVTSRRDGDSPSHGVCKPKLIINKNSRKSKSPIPGQNYLIGTTERPSVSVSH SQSDYHRQTPGMAWDRPSTGSGTQPPYFNLTIRMLSTPTPIACAPSAITQQAAPHQ QSRWIEREPALLSAQVETLSDD"							
ORIGIN								
Alignment Scores:								
Pred. No.:	3,74e-07	Length:	1599					
Score:	714.50	Matches:	204					
Percent Similarity:	51.22%	Conservative:	68					
Best Local Similarity:	38.42%	Mismatches:	194					
Query Match:	5.41%	Indels:	65					
DB:	10	Gaps:	23					
US-09-522-753-5 (1-2517) x AF124821 (1-1599)								
Qy	2016	AlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGlu	2035					
Db	112	GCCTCGGCAAGATGCAAGGGACGGTGGCTCTCAAAAGTTTCAGACTTCTCT	162					
Qy	2036	LeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro	2055					
Db	163	-----AGTAGCTGTCTTCTCACAGGTATGACGACCTAGTAGTGATGTAGGTG	213					
Qy	2056	ValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGlu	2075					

Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics -----
Center Project name: H_NH1399P15

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Taten, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

Uncertain bases from base 63497 to base 63498.

Uncertain bases from base 186010 to base 186011.

The sequence from base 942 to base 1018 was derived from one plasmid subclone.

The sequence from base 134927 to base 135119 was derived from one plasmid subclone.

There is an unresolved tandem repeat from base 1359 to 12000.

This sequence is the entire insert of the clone.

FEATURES	
source	Location/Qualifiers
	1..192181
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/map="2"
	/chromosome="2"
	/clone="RP11-1399P15"
	/clone_lib="RPCI-11"
repeat_region	2..100
	/rpt_family="Alu"
repeat_region	101..160
	/rpt_family="Alu"
repeat_region	161..376
	/rpt_family="Alu"
repeat_region	377..436
	/rpt_family="Alu"
repeat_region	437..647
	/rpt_family="Alu"
repeat_region	648..702
	/rpt_family="Alu"
repeat_region	703..919
	/rpt_family="Alu"
repeat_region	920..970

QY	2413	LySaLaYsSerProAlaProGly-----LeuAlaSerGlyAspArgProSerVal	2430
Db	1279	AGTCTAAATCTCTTCTGGCAGAACTACTTAGGAAGCTTCTCTCTGTC	1338
QY	2431	SerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrp	2450
Db	1339	TCCTCTGTACATTCAGAGGGTGATTACACAGCGACACACCA-----GGATGGCGCTTGG	1392
QY	2451	GlusAspArgProSerSerAlaGlySerThrProPheProTyAsnProLeuIleMetArg	2470
Db	1393	GAAGACAGCGCGCTTCTCAACAGCTCCATCAGTTCCTTACACCTCTGACCATCCGG	1452
QY	2471	LeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGly----	2489
Db	1453	ATG-----CTGACGACTACCCCGCAACCCGATCGCTGTGCCCCGTCAGGCCATC	1503
QY	2490	ProLeuAlaGlyProHis-----HisAlaTrpAspGluGluProLysProLeuLeu	2506
Db	1504	ACCGAGCAGCTCCGATCAACAGCGCGCATCTGGAGCGGAGCGCCCGCTCCCTC	1563
QY	2507	CysSerGlnTrpGluThrLeuSerAspSerGlu	2517
Db	1564	TCAGCACAGTATGAGACACTGTCTGACAGTGAC	1596

RESULT 62	
AC133644	192181 bp DNA linear PRI 23-NOV-2002
LOCUS	
DEFINITION	Homo sapiens BAC clone RP11-1399P15 from 2, complete sequence.
ACCESSION	AC133644
VERSION	AC133644.3 GI:25141136
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 192181)
TITLE	Sulston, J.E. and Waterston, R.
JOURNAL	Toward a complete human genome sequence
MEDLINE	Genome Res. 8 (11), 1097-1108 (1998)
PUBMED	99063792
REFERENCE	2 (bases 1 to 192181)
AUTHORS	Kim, K., Bielicki, L. and Creason, K.
TITLE	The sequence of Homo sapiens BAC clone RP11-1399P15
JOURNAL	Unpublished (2001)
REFERENCE	3 (bases 1 to 192181)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (16-SEP-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	4 (bases 1 to 192181)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (13-NOV-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	5 (bases 1 to 192181)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (21-NOV-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	6 (bases 1 to 192181)
AUTHORS	Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (23-NOV-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT	On Nov 21, 2002 this sequence version replaced gi:24943062. ----- Genome Center Center: Washington University Genome Sequencing Center Center code: WUGSC

repeat_region	18335..18485	/rpt_family="Alu"
repeat_region	18820..18918	/rpt_family="Alu"
repeat_region	19570..19740	/rpt_family="L2"
repeat_region	19741..20036	/rpt_family="L1"
repeat_region	19741..20036	/rpt_family="Alu"
Alignment Scores:		
Pred. No.:	2.22e-05	Length: 192181
Score:	709.50	Matches: 596
Percent Similarity:	33.88%	Conservative: 351
Best Local Similarity:	21.32%	Mismatches: 1186
Query Match:	5.37%	Indels: 674
DB:	9	Gaps: 115
US-09-522-753-5 (1-2517) x AC133644 (1-192181)		
Qy	15	ThrGluProArgTyrProHisSerLeuSerTyrProValGlnIleAlaArgThrHis 34
Db	3454	AGCCAGCCCAAGTGGCCCAAGCCACCTAGTCAAG---CCAGCTAAGCCAGCCAGCCAGCCCA 3510
Qy	35	ThrAspValGlyLeuLeuGluTyrGlnHisSerArgAspTyrAlaSerHisLeuSer 54
Db	3511	AGCCAGCCCAAGCCAGCCCAAGCCCTCCAGCCAGCCCAAGCAAGCAAGCATGCCAGCCAGCCAGC 3570
Qy	55	ProGlySerIleIleGlnProGlnArgArgProSerLeuSerGluPheGlnPro 74
Db	3571	CAGGGAGCC-----AAGCCAGCCAGCAAGCAAGCCAGCCAGCCAGCCAGCCAGCCCA 3624
Qy	75	GlyAsnGluArgSerGlnGluLeuHis-----LeuArgPro 86
Db	3625	GGGAGCCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCCA 3684
Qy	87	GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLys 106
Db	3685	AGCCAGCCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCT 3744
Qy	107	ArgProArgLeuGluLeuLeuPro-----AspProLeuLeuArgProSer--- 121
Db	3745	AACCCAGCCAG-CAAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3803
Qy	122	-----ProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAsp 138
Db	3804	CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 3839
Qy	139	ArgSerLeuThrGlyLysLeuGluProValSerProProSerProHisThrAspPro 158
Db	3840	-----AAACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCCA 3878
Qy	159	GluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArg 178
Db	3879	AGC-----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3908
Qy	179	ValAspArgGluIleThrMetValGluGlnGlnIle----- 190
Db	3909	-----ACCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGTCA 3953
Qy	191	-----Ser 191
Db	3954	GCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCA 4013
Qy	192	LysLeuLysLysLysGlnGlnLeuGluGluAlaAlaLysProGluProGlu 211
Db	4014	GCCTGCCAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 4073
Qy	212	LysProValSerProProIleGluSerLysHisArgSerLeuValGlnIleTyr 231
Db	4074	AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 4103
Qy	232	AspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGln 251

Qy 937 euLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerP 957
Db TAGCCAAAGCCA-----GCCAAGCCACCGCGCAGCCCAAGCCAGCCAGCCAAACC 5987
Qy 957 roGlnLysProLeuAsp-----LeuLysGlnLeuLysGlnArgAlaAla- 971
Db CAGCCCAAGCCAGCCAAAGCAGCAGACATTAGCTAGCCAGCCAGCCAGCCAGCCAGCC 6047
Qy 972 -----AlaIleProProIleGlnValThrLysValHisG 983
Db AGTCAAGCCGCGCAGCTGCCAGCCCAACCAAGCCACCGCGCAGCCAGCCAGCCAGCC 6107
Qy 983 luPro- ProArgGluAspAlaPro-----ThrLysProAla 995
Db AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6167
Qy 996 ProProAlaProProProGlnAsnLeuGlnProGluSerAspAla-----ProGln 1013
Db AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6227
Qy 1014 GlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGlu 1033
Db CAGCCA---AGCCACCAGCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6281
Qy 1034 AlaPheAlaAlaGlnLysLeuProGlyAspProProCysTrpThrSerGlyLeu 1053
Db -----GCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6335
Qy 1054 ProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspSer 1073
Db AGCCAAACAGCCAAAGCCA-----GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6383
Qy 1074 AlaPheSerTyraProProGlyHisProLeuProLeuGlyLeuHisAspThr----- 1091
Db GCCAGCCACTCAAGCCATCCAGCAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6443
Qy 1092 AlaArgProValLeuProArgProProThrIle-----SerAsnProProLeuIle 1109
Db GCCAAGCCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6503
Qy 1110 SerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyWet 1129
Db AGCCAAAGCCGCGCGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6563
Qy 1130 SerValGlnLeuHisValProTySerGluHisAlaLysAlaProValGlyProValThr 1149
Db CAC-----CCGCGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6605
Qy 1150 MetGlyLeuProLeuProMetAsp-----ProLysLysLeuAlaProPhe 1164
Db ACAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6665
Qy 1165 SerGlyValLysGlnGluLeuSerProArgGlyGlnAlaGlyProProGluSerLeu 1184
Db GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6725
Qy 1185 GlyValProThrAla-----GlnGluAlaSerValLeuArgGlyThrAlaLeuGly 1201
Db AGCCACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6785
Qy 1202 SerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSer 1221
Db GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6845
Qy 1222 AlaIleThrTyraArgGlySerIleThrHisGlyThrProAlaAspValLeuTyLysGly 1241
Db GCC----- 6848
Qy 1242 ThrIleThrArgIleIleGlyGluAspSerProSerArg---LeuAspArgGlyArgGlu 1260
Db -----AACCAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6884
Qy 1261 AspSerLeuProLysGlyHisValIleTyraGluGlyLysLysGlyHisValLeuSerTy 1280

Db 6885 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6911
Qy 1281 GluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyPro 1300
Db -----CAGCAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 6944
Qy 1301 ProHisGluThrAlaAlaProLysArgThrTyraAspMetMetGluGlyArgValGlyArg 1320
Db -----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7001
Qy 1321 AlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
Db GCCATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7049
Qy 1341 SerProHisHisLeuLysGlu----- 1347
Db TCACCCCACTCAAACTAGCCCAAGTCACCGCGCAGCCAGCCAGCCAGCCAGCCAGCCAG 7109
Qy 1348 -----GlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSer 1362
Db CCCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7168
Qy 1363 TyrValGluAlaGlnGluAspTyraLeuArgArgGluAlaLysLeuLysArgGluGly 1382
Db AGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7219
Qy 1383 ThrProProProProProProSerArgAspLeuThrGluAla---TyrLysThrGlnAla 1401
Db -----CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7273
Qy 1402 LeuGlyProLeuLysLeu-----LysProAlaHisGluGlyLeuValAlaThr---Val 1418
Db ATCAAAACAGCCCAAGTCACCGCGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7333
Qy 1419 LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGlu 1438
Db AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7393
Qy 1439 LeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLys 1458
Db AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7453
Qy 1459 TyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuIleGly 1478
Db CCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7501
Qy 1479 SerProGlyArg-----ThrPheProProValHisPro 1489
Db CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7561
Qy 1490 LeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyraGluGluSerLeuLys 1509
Db GG-CCAGCCCTGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7620
Qy 1510 SerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaProValIle 1529
Db AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7671
Qy 1530 ValProGluLeuGlyLysProArgGln-----SerProLeuThrTyraGluAspHisGly 1547
Db -----GGCCAGCAGCAGCAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7716
Qy 1548 AlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThrPro 1567
Db -----CCAAGA---GAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7751
Qy 1568 ArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSer 1587
Db ----- 7752
Qy 7752 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 7799
Qy 1588 ThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHisPro 1607

[illegible][illegible]


```
Db 173700 GTTGGCTCCGGTCCCGGGGCGCGCTTGGCCGATCAGCGGGCGCCGGTGGC 173641
Qy 1339 ArgHisSerProHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGly 1358
Db 173640 CGCACGAATTGGCGTGTGATCGGGGCCAGCAGCGCGACACGGC-----GGC 173593
Qy 1359 IleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeu 1378
Db 173592 GG-CCTCG-----CGCGCATACGGCGCTCCGC 173564
Qy 1379 LysArgGluGly-----ThrProPro-----ProProPro 1389
Db 173563 CGCGCTCAGGGCGTGCACGAATGGCGGTAAACCGCGCGCTGTGCGCTCGCGCGCT 173504
Qy 1390 SerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGly-----ProLeu 1405
Db 173503 GTTAG-----CCTGGCTTGGGCACCGAACACCGCGCAA 173468
Qy 1406 LysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIle 1425
Db 173467 TCGTGTGACACCTCATCGGCACCGCGGTCAACAATCCCGTGTGACGCGCGCG 173408
Qy 1426 HisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgPro 1445
Db 173407 CCGCATTAGCGAGCTAATCGCGACCGCGATCCCGGCAAAATCGGTGCGCGCGCACCA 173348
Qy 1446 LeuLysGluGly-----SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSer 1464
Db 173347 CC-AGCTCCGGCGAACCATCACC-----173325
Qy 1465 ThrThrGlySerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPhe 1484
Db 173324 -----CATGATCGCATGCTC-----173307
Qy 1485 ProProValHisProLeu-----AspValMetAlaAspAlaLeuGluArgAla 1502
Db 173306 CCAACCCCTCAGCGCTGTGTCGCGATACAGCGACACACCCGTAGC-----GGTAGAAT 173250
Qy 1503 CysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSer-----1517
Db 173249 TAATTGCAGAAATCAACGATATCGGATCCGGCACCGCATGACGCTTCACACGTTTC 173190
Qy 1518 SerGlyGlySer-----IleAlaArgGlyAlaProValIleValProGluLeuGlyLys 1535
Db 173189 GCAGCGCGCAGCTGTGACGGCTACGGGTGCGG-CGGCGTGTGCCAGAACCGGCGG 173131
Qy 1536 ProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPro 1555
Db 173130 CGGTCCCGGCA-----CAGGGCGCGCGCAACCGGCAC-----173095
Qy 1556 ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSer 1575
Db 173094 -----CGCGCGGATACCGTGTGCACGACGACCGCGA-----CGGGCC 173056
Qy 1576 SerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSer 1595
Db 173055 TCTGAAAGCGCTCGAGCGCGG-----GGAICTTCGCCAGATACCTGCGCGCGCGG 173002
Qy 1596 Pro-----1596
Db 173001 CCCCATTTTGGGCGCGCTCCGCGCGCATCTGCATCAGCGCGAGTGTGACGCGC 172942
Qy 1597 ---HisSerThrValProGluHisHisProHisProHisProTyrTyrGluHisLeu 1615
Db 172941 GGAAGCCATCGCGTCCCGCGCTACGGCAACCGTTGAGCCCG-----172897
Qy 1616 ArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThr 1635
Db 172896 -----TTTGGCCACG 172885
Qy 1636 SerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeu 1655
|||
|||
|||
```

```
Db 172884 AGCAACCCCGCGCGCGCGCTTCCGGGGTCCGCGCG-----TTG 172843
Qy 1656 AlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAsp 1675
Db 172842 GCGCGGTGCGCGCGCTTCCCGCGGTGCGCGGTGCGATCAGCCCGCGCGTCCGCGG 172783
Qy 1676 ThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMet 1695
Db 172782 AGCCACCGCGCGCG-----172768
Qy 1696 HisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro 1715
Db 172767 -----CCGGTGTGCGCG 172756
Qy 1716 ArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSer 1735
Db 172755 AGCTTAAGGTGCGCGCTTCCGCGCGCGCGCGCG-----172717
Qy 1736 GlnValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMet 1755
Db 172716 ---CTACCCCAATACTGACCGCGGTGCGCGGTTCACCGCGCGCGCGGTAGCGTTA 172660
Qy 1756 AspArgLeuAlaTyrIleuProThrAlaProGlnProPheSer-SerArgHisSerSerSe 1775
Db 172659 CCGCGATTCGCGCATGTCGCCCGCGCGCGCGAAGCCAGCCAGCGCATGTCGCGCG 172600
Qy 1775 rProLeuSerPro-----GlyGlyProThrHisLeuThrLysProThrTh 1790
Db 172599 GCGCCACCGACCCCGCGCGGTGCGGAACCGCGCAACCGCGCGCGCGCGCGTGGAC 172540
Qy 1790 rThrSerSerSerGluArgGluArgAspArgAspArgGluArgAsp-----ArgAs 1807
Db 172539 AACAGCAGCGCGCGGTGCGCACCGCGCGCTCCACCGAGAAACCAATCCGCGG 172480
Qy 1807 pArgGluArgGluLysSerIleLeuThrSerThrThrThrValGlu-----HisAlaPr 1825
Db 172479 GCGCGCGCGCGCGCGCTGCGCAACAGCAGCCCTCGCGCTACCGCGCGCGCGCGCTCC 172420
Qy 1825 oIleTpArgProGlyThrGluGln-SerSerGlySerSerGlySerSer-----GlyG 1843
Db 172419 CMACCGGTGTCAGGTTTCGCGGTGCGGTACCGCGGTCCCGCGCGCGCGCGCGCGCGG 172360
Qy 1843 lyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProI 1863
Db 172359 AGCATTCGCGCATTCGCGCGCGCGCGCAACCGCGCGCTAGTGTGTAGCCCG 172300
Qy 1863 leSerProArgThrGlnAspAlaLeuGlnAlaProSerValLeuHisAsnThrGlyM 1883
Db 172299 CCGGTCCCGCGTACCTCCGCGCGCACCGATCAGCCCGCG-----CAGCAGTCCGCGGTC 172243
Qy 1883 etLysGlyIleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSert 1903
Db 172242 CCGCGCGC-----CCACCGCGCGCGCGCTC-----GCCA 172213
Qy 1903 hrSerSer-ProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGly 1922
Db 172212 CCAACAAGCGCGTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 172159
Qy 1923 GlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGluAla 1942
Db 172158 -----CCGACCCCGCTCCCGCGCGCGCGCGCGT-----172126
Qy 1943 ProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLys 1962
Db 172125 ---ACGGTGTGCTCGCGCATCCCGCGCGC-----172096
Qy 1963 ProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArg 1982
Db 172095 CCGCGCGTCG-----CGAGCAGCGCAGCGCG-----CCGGCC 172066
Qy 1983 ProLeuValProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeu 2002
|||
|||
|||
Db 172065 CCGCGCGCGCGCGCGGTACCC---CACAGCGCG---GCAGCCCGCGCGCGCGCGCGG 172012
```



```

REFERENCE
AUTHORS
3 (bases 1 to 200389)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Maclean,C., Macdonald,P., Major,J.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 200389)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Maclean,C., Macdonald,P., Major,J.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 16, 2003 this sequence version replaced gi:26449133.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L28245
Center clone name: 199_M_12
-----
Location/Qualifiers
1..200389
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP13-199M12"
/clone_lib="RP13-13 Human Female BAC"
1881..2390
/rpt_family="LOR1a"
complement(2549..2616)
/rpt_family="MIR"
complement(2782..2902)
/rpt_family="L2"
complement(3047..3272)

repeat_region
repeat_region
repeat_region
repeat_region

FEATURES
source
1..200389
/rpt_family="MIR"
complement(3553..3716)
/rpt_family="MIR"
complement(3849..3912)
/rpt_family="MLTID"
3913..4179
/rpt_family="MLTID"
/rpt_family="AluSc"
complement(4180..4615)
/rpt_family="MLTID"
complement(4630..4708)
/rpt_family="L2"
complement(4725..4802)
/rpt_family="L2"
complement(5008..6173)
/rpt_family="L1PA3"
7787..7914
/rpt_family="L2"
8285..8348
/rpt_family="AluY"
8569..8611
/rpt_family="CT-rich"
8612..8642
/rpt_family="(CA)n"
8657..8791
/rpt_family="(CA)n"
8827..8878
/rpt_family="(CA)n"
9186..9484
/rpt_family="AluSx"
complement(9842..9983)
/rpt_family="L2"
10727..10981
/rpt_family="MIR"
11251..11705
/rpt_family="MLT2B1"
complement(11833..12418)
/rpt_family="MER82"
complement(12948..13100)
/rpt_family="LIM4"
complement(13219..14843)
/rpt_family="LIPAB"
14843..15616
/rpt_family="LIP"
complement(15907..16160)
/rpt_family="HERV16"
complement(16198..16311)
/rpt_family="HERV16"
complement(16319..16620)
/rpt_family="AluY"
complement(17103..17204)
/rpt_family="MER20"
17668..17813
/rpt_family="MIR"
17838..17880
/rpt_family="(TA)n"
18877..19017
/rpt_family="MIR3"
19402..19424
/rpt_family="AT-rich"
19638..19658
/rpt_family="AT-rich"
20031..20100
/rpt_family="L2"
20638..20704
/rpt_family="AT-rich"
complement(21399..21823)
/rpt_family="L1MB4"
complement(21829..22010)
/rpt_family="L1PA10"
22011..22212
/rpt_family="L2"
complement(23550..25788)
/rpt_family="L1MD2"

```

```

repeat_region complement(25789..26210)
repeat_region /rpt family="MSTA"
repeat_region complement(26211..27822)
repeat_region /rpt family="LiMD2"
repeat_region 28169..28241
repeat_region /rpt family="MSTD"
repeat_region complement(28242..28408)
repeat_region /rpt family="LiMD2"
repeat_region complement(28418..28725)
repeat_region /rpt family="LiMD2"
repeat_region 31611..31646
repeat_region /rpt family="GA-rich"
repeat_region 33386..33420
repeat_region /rpt family="AluDb"
repeat_region 33421..33459
repeat_region /rpt family="(CA)n"
repeat_region 33460..33615
repeat_region /rpt family="AluDb"
repeat_region 33616..33929

Alignment Scores:
Pred. No.: 2,41e-05 Length: 200389
Score: 708.50 Matches: 569
Percent Similarity: 32.94% Conservative: 328
Best Local Similarity: 20.90% Mismatches: 1195
Query Match: 5.36% Indels: 641
DB: 9 Gaps: 102

US-09-522-753-5 (1-2517) x AC134678 (1-200389)
Qy 2 SerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProPro 21
Db 52836 GCTGGCAAGACAGCCAGCCACTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 52777
Qy 22 HisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeuGlu 41
Db 52776 AGCCAGCAAGCCACTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 52717
Qy 42 TyrGln-HisHisSerArgAspTyr-----AlaSerHisLeuSerProGlySerIleI 59
Db 52716 TTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 52662
Qy 59 eGlnPro-----GlnArgArgProSer-----LeuLeuSerGluPheG 73
Db 52661 -CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 52603
Qy 73 nProGlyAsnGluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPr 93
Db 52602 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 52552
Qy 93 oGluLeuGlyLysSerGluMetGluPheIleGluSerLys-ArgProArgLeuGluLeuL 113
Db 52551 AGCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 52492
Qy 113 eu-----ProAsp-----ProLeuLeuArgProSer- 121
Db 52491 CCCAAGCGGCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 52432
Qy 122 -----ProLeuLeuAlaThrGlyGlnProAla---GlySerGluAspLeuT 136
Db 52431 GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 52372
Qy 136 hrLysAspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHist 156
Db 52371 GCCAAGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 52312
Qy 156 hrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuIleGlnAsnM 176
Db 52311 GCCAACCCTGACGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 52252
Qy 176 etAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysL 196
Db 52251 CAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 52193

```

```

Qy 196 ysGlnGlnLeuGluGluAlaAlaLysProGluProGluProGluLysProValSerP 216
Db 52192 -----ACAGCCAAAGCCACT-CAGCCGCGCCAAAGCCAGCCAGCCAGCC 52157
Qy 216 ro-----ProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluA 234
Db 52156 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 52127
Qy 234 snArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluL 254
Db 52126 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 52091
Qy 254 euProLeuTyrAsnGlnProSerAspThrArgGln-----TyrHisGluA 269
Db 52090 AGCCACCTTC- -CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 52034
Qy 269 snIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnH 289
Db 52033 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 51974
Qy 289 isAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaL 309
Db 51973 ACCAGCCAGCCAG-----ACAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51960
Qy 309 euGluLysLysValGluArgIleGluAsnAsnProArgArgAlaLysGluSerLysV 329
Db 51959 -----ACAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51924
Qy 329 alArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluA 349
Db 51923 -----GAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51878
Qy 349 rgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerG 369
Db 51877 CCGGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51818
Qy 369 luHisGluValSerGluIleIleAspGlyLeuSerGluGlnGlnLeuGluLysGlnM 389
Db 51817 GCCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51758
Qy 389 etArgGlnLeu-----AlaValIleProProMet-----L 399
Db 51757 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 51698
Qy 399 euTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspP 419
Db 51697 GCCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51671
Qy 419 roMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluT 439
Db 51670 CAGTCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 51612
Qy 439 hrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuG 459
Db 51611 -----AGACGCTAGCCAGTCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51560
Qy 459 luArgLysThrValAlaGluCysValLeuTyrTyrTyrTyrLeuThrLysLysAsnGluAsnT 479
Db 51559 AAAAGCCAGCC----- 51549
Qy 479 yrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnG 499
Db 51548 --AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51491
Qy 499 lnglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 513
Db 51490 CCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51431
Qy 514 -----ArgSerSerGlnGluGluL 520
Db 51430 GGCCACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 51371

```


QY	1876	erValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValAlaGluProSerLysProT	1896
DB	47527	CC-----AAGCAGCGCAAGCAAGCCACCAGCAGCAAGCCAGCA--	47490
QY	1896	hrValLeuArgSerThrSerSerProValArgProAlaAlaThrPheProProA	1916
DB	47489	-----ACCAGAG-AGCCAGGCAATTCAGCGCTAGACAGCCAGCAAGCCAGCAG	47445
QY	1916	laThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProV	1936
DB	47444	CCAGTCATGCC-----	47434
QY	1936	alLeuLeuProLysGluAlaProArgValAlaArgProGluUArgProArgAlaAspThrG	1956
DB	47433	-----AGCCAGCAGCCCAAGTCAGCAGCCAGCCAGCAAGCA--	47398
QY	1956	lyHisAlaPheLeuAlaLysPro-ProAlaArgSerGlyLeuGluProAlaSerSerPro	1975
DB	47397	-----GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47352
QY	1976	SerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAla	1995
DB	47351	AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47295
QY	1996	ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProPro	2015
DB	47294	ACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47247
QY	2016	AlaSer--AlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGln	2034
DB	47246	GCAGCCCAAGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47190
QY	2035	GluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGlu	2054
DB	47189	CCAGCCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47130
QY	2055	ProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeu	2074
DB	47129	CCACTGAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47103
QY	2075	GluGluLeuAspLysSerHisLeuGluGlyLeuArgProLysGlnProGly--Pro	2093
DB	47102	-----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47079
QY	2094	ValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProGluSerGln	2113
DB	47078	ACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	47025
QY	2114	ProSerSerPro-LeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValVa	2133
DB	47024	CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46973
QY	2133	lThrLeuAlaGlnHisLysSerGluValIleThrGlnAspTyrThrArgHisHisPro--	2152
DB	46972	-----GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46929
QY	2153	-----GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPh	2166
DB	46928	CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46869
QY	2166	eProGlyAlaSerCysProValLeuAspLeuArgProProSerAspLeuTyrLeuPr	2186
DB	46868	CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46821
QY	2186	oProProAspHisGlyAlaProAlaArgLysProHisSerGluGlyLysArgSe	2206
DB	46820	AAG-CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46762
QY	2206	rProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSe	2226
DB	46761	CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	46729
QY	2226	r-----ProProGluGlyMetThrGluProGlyHisHisSerArgSe	2239

[illegible]

RESULT 65	SC0939111/c	321250 bp	DNA	linear	BCT 11-FEB-2003
LOCUS	SC0939111				
DEFINITION	Streptomyces coelicolor A3(2) complete genome; segment 8/29.				
ACCESSION	AL939111	AL049497	AL096872	AL109684	AL109732
	AL109849	AL113219	AL113220	AL113650	AL1355740
	AL590743	AL592262	AL645882		AL445343
VERSION	AL939111.1	GI:24419015			
KEYWORDS	Streptomyces coelicolor A3(2)				
SOURCE					

QY 1035 PheAlaAlaGluAlaGlnLysLeuProGlyAspProCysTrpThrSer-----Gly 1052
Db 61269 -----CGTGCCTGGCGGACCGGACCTGCGAGTCTGGTGGC 61228
QY 1053 LeuProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspPro 1072
Db 61227 GCATCTCACCGGCTCGCGTCTCGCGGCCACCGACGAGGCGCATCGCGCGGAGC 61168
QY 1073 SerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAla 1092
Db 61167 TGGCGAACGACCGAGCGCACCGCGCAGGAGGTCGCGCGCTGCGCGCGCGCTGC 61108
QY 1093 ArgProValLeuProArgPro-----Pro 1100
Db 61107 CGGACGCGGAGCGCAAGCGCGCGCTGGGAGGCGATGTCGCTCGGACGACTCTCCA 61048
QY 1101 ThrIleSerAsnProProProLeuIleSerSerAlaLysHisPro-SerValLeuGluAr 1120
Db 61047 ACTACCTGTTCACGCGCACCG-----CCGAGGCTTCTGGCAGC 61009
QY 1120 GlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHi 1140
Db 61008 CCGAACAGGCGGAGCTGTCGCGGACTACGTACCGCGCTA---CTACCCCGAGCGCGTCG 60952
QY 1140 sAlaLysAlaProValGlyProValThr-----MetGlyLe 1152
Db 60951 CGTCTGGCGGCGG-----CGCGGTCCCGCCATGGCCGACGCGCGCGCTGGGCGCT 60898
QY 1152 uProLeuProMetAspProLysLys---LeuAlaProPheSerGlyValLysGlnGlu1 1171
Db 60897 TCCCGCGCCACGCGGTGCGACCGGACACCTCGCGCTGGCGGAGTGCCTCGCGGACG 60838
QY 1171 nLeuSer-----ProArgGlyGlnAlaGlyProProGluSerLeuGlyValProTh 1188
Db 60837 CCGATCCGATCCCGCGCTCGCGCGCAAGCTCACCGACAGCTCGACGA-----CCTGG 60784
QY 1188 rAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGly-----1201
Db 60783 CGCGCGCTTCGGGTAG-----GGAGCGGACACGAGTACCCGCGCGCGCGCGCG 60730
QY 1202 -SerValProGlyGlySerIleThrLysGlyIleProSerThrArgValProSerAspSe 1221
Db 60729 CCCGCGCGCGCGCGGCGC-----60711
QY 1221 rAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysG1 1241
Db 60710 -----GGCCGTGCCCCCGTGGGGAGCGCGCGGATACCCCTTTCGGGT 60667
QY 1241 yThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAs 1261
Db 60666 TGTGATCAGTGTGTCGGGGGTACACGCCCGCGCACGTGGAAGCTGGAACCCAGCT 60607
QY 1261 pSerLeuProLysGlyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrG1 1281
Db 60606 CCCGTCGCCGAGGACAGCTCATG-----60582
QY 1281 uGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProPr 1301
Db 60581 -----GGCATGCC-GCCCTCGCTCGGCGCGCGCACGCGCGCGCGCTCGCGCCCTGC 60527
QY 1301 o---HisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyAr 1320
Db 60526 TCACACCGTGTCTGACGCGCTCGCGCGCGGAC-----GGCGCGACGCGGCG 60479
QY 1320 gAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHi 1340
Db 60478 GCCGCTGCC-----CGCGCGCGCGCGCGCGCGCTCG 60446
QY 1340 sSerProHisHisLeuLys---GluGlnHisHisIleArgGlySerIleThrGlnGly11 1359
Db 60445 CCGCCCGGTGGCGGACGCGCGCGCGCGCCATCTGCGCGGAGAGGCGCGACCCGAGGCGC 60386

QY 1359 eProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLy 1379
Db 60385 TCCGCGTCTCTGT-----CACCGCGCTGGCGGAGGCGCGC 60350
QY 1379 sArgGluGlyThrPro-----ProProProProProSer-----1390
Db 60349 CCGACCGCGCCACCCCTGTGCGCGCGCCACCTGCACACCCCGCGCTCGCGTGGCGG 60290
QY 1391 -ArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
Db 60289 CCGCGCCCATCTTCGCGCGGAGGTCTCAACCCCTC---CCTCGACTCTCGGACAGG 60233
QY 1410 aHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIlePro-- 1429
Db 60232 CACCCGCGCGCTC-----CGCGTGGAGCCCTCGTCAACCGCGCGCTGG 60188
QY 1430 -----ArgGluGluLeuArgHisThrProGluLeuProLeuAlaProAr 1444
Db 60187 CCGCGGAGACCGCGCGCGCGCGCTGTCAACCA-CGCGAG-----GCACCGAG 60138
QY 1444 gProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSe 1464
Db 60137 TCAACACAGCTCGCGTCTGCTCGCGCGGAGCGGACGCGCGCGGCTGGCGGTGTC 60078
QY 1464 rThrThrGlySerLysLysHisAspValArgSer-----LeuIleGlySerProG1 1481
Db 60077 CAGGGG-CGCGGCGCCACCATCTGCTGCCCGCGCGCTGCTGCTCGCGCTGCC-- 60021
QY 1481 yArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluAr 1501
Db 60020 -----GAGCGCGTCTGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 59980
QY 1501 gAlaCysTyrGluGluSerLeuLysSerArgProGly-----ThrAl 1515
Db 59979 CGCGCGCTGGAGAGGCGCTCACCCAGGTCCCGCGCGCGCGCGCGCGCGCGCGCG 59920
QY 1515 aSerSerSerGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLy 1535
Db 59919 GGGCACACCGACCGCGCGGCTCATCGACCGCTGCCGAGATCGCGCA-CGCTGGCGCG 59861
QY 1535 sProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuPr 1555
Db 59860 CCGACGCGCGCGCGCT-----CCACGTCGACGCGCGCGCTACG 59825
QY 1555 oArgGlySerProValThrMetArgGluProThrPro---ArgLeuGlnGluGlySerLe 1574
Db 59824 GCGCGGCGCTCTGTTCAGGAGCGGACCGCGCGCGCGCTCGCGGAGTGGAGCGCGCG 59765
QY 1574 uSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLy 1594
Db 59764 ACACCGTCGCCCTGGACCTGCACAGCTCGGCTGGCAGCCCATCCCGC-----CG 59714
QY 1594 sSerProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLe 1614
Db 59713 GACTCCTCAAGT-----CACCGACGCGCGCGCGCGCTGG-CGCGCC-----59675
QY 1614 uLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspPr 1634
Db 59674 -----TCCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 59646
QY 1634 oThrSerIlePro---ArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProAr 1653
Db 59645 GACGACGACACCGACGCGGCGCTGCCGACCTCTCTCGGGCGCT-----CGCGG 59598
QY 1653 gHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTy 1673
Db 59597 CGCACGAGCGCGCGCGCGCGCGCTCTGAAGACCGCGCTCACCTCAAGACCCCTGGGCGC 59538
QY 1673 rProAspThrAlaAlaLeuGluAsnArgGlnThrIleIle-----AsnAspTyrIleTh 1691
Db 59537 GCGGAGTGGCGCGCGCTCTGTCGAGCGGCTGTCTCTCGCGCGCGGAGTTCGCGGATCTG 59478
QY 1691 rSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuAr 1711


```
Db 59477 GTCCAGACCCATCCCGGGTTTCGAATGTCACACCGCCCGCCACCATCAGCACGGTCTCTGTC 59418
Qy |||::: |||
Db 1711 gGlyLeuSerProArgGlu-----SerSerLe 1720
Db 59417 CGGCCCGCGGAGCGGAGACGACCGCGTGCCTGCGCGCGTACCGCGCGGCTGCTCACCACC 59358
Qy |||::: |||
Db 1720 uAlaLeuAsnTyrAlaAla-GlyProArgGlyIleAlaAspLeuSerGln----- 1736
Qy |||::: |||
Db 59357 GCGCGCGCGTACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 59298
Qy |||::: |||
Db 1737 -----ValProHisLeuProValLeuValProProThrProGlyThrProA 1752
Db 59297 CTGAACCGCGCACCCCGCCAGACGACCTCGCAGCGCTCTCTGAGCGCTGTGGAGAGACAC 59238
Qy |||::: |||
Db 1752 laThrAlaMetAspArgLeuAla-TyrLeuProThrAlaProGlnProPheSerSerArg 1771
Db 59237 GTCCCGGATGAACACCCCGGACCCGACACCCCGGAGCGCCCGCGGACCTCTCGGCA 59178
Qy |||::: |||
Db 1772 HisSerSerPro---LeuSerProGlyGlyProThrHisLeuThrLysProThrThr 1790
Db 59177 TCGGCATCGGCCCTTCAACCTCTCCCTCGCGGCGCTCGCGACCCCTCACCGCGCTCG 59118
Qy |||::: |||
Db 1791 ThrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArg 1810
Db 59117 ACGCCGTCTTACGACACCGCGCCCG----- 59091
Qy |||::: |||
Db 1811 GluLysSerIleLeuThrSerThrThrValGluHisAla----- 1824
Db 59090 GCTTCCCGTGCACCCCGCGCTGTCTATCGACGCGCAGCGCTCCAGGTGCGCTTCCTCG 59031
Qy |||::: |||
Db 1825 ProfileTrpArgPro---GlyThrGluLysSerSerGlySerSerGlyGly 1843
Db 59030 CCGACCTGTGCACCTCTCGTACCGCGCGGACCCCTGGACCTTCTCAACACCTCAGGG 58971
Qy |||::: |||
Db 1844 GlyGlyGlySerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIle 1863
Db 58970 CCAGAGAAGCGCTCTCCCGTTCTACTTCGCGGAGCGCTTCCACATCAGCGCGCGAGT 58911
Qy |||::: |||
Db 1864 -----SerProArgThr----- 1867
Db 58910 ACGAGCGCTACTCGCGTGGTTCGCGGAGAACCTTCGCGCGCTCCGCTTCCACACCGAG 58851
Qy |||::: |||
Db 1867 ----- 1867
Db 58850 TCGAGCGCGTGGTGGNACCCGACGCGACCTTTCGAGTGCATACACCGAGTGC 58791
Qy |||::: |||
Db 1868 GlnAspAlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIle 1887
Db 58790 ACGCGCGGCGGAGCGCGGAGCGCGCTCGCGCGCACCCACACCGAGAACTGCTCTCGGCA 58731
Qy |||::: |||
Db 1888 ThrAlaValGluProSerLysProThrValLeuArgSerThrSerSer----- 1904
Db 58730 TCGGCACCGAACACACAGCTCCCGGAC-----CGCTCAGACCCCTCGTCGAGGACCCCG 58677
Qy |||::: |||
Db 1905 ---SerProValArgProAlaAlaThrPhe-ProProAlaThrHisCysPro----- 1920
Db 58676 CGGTCCCGTGTCTCCACCGCGCGGACTACCTCCGCGCACCGGACACGCTCTCTGCGCGCG 58617
Qy |||::: |||
Db 1921 -----LeuGlyGlyThrLeuAspGlyValTy 1929
Db 58616 GACAGTCCAGCTCATCGGCACCGCGCGAGTCCGCGCGGAGGT----- 58574
Qy |||::: |||
Db 1929 rProThrLeuMetGluProValLeuLeuProLysGluAlaPro----- 1943
Db 58573 -----CTTCTCGAGCTGCTCCGCCACCGCGCCCGCGGAC 58539
Qy |||::: |||
Db 1944 -----ArgValAlaArgPro---GluArgProArgAlaAspThrGlyHisAlaPh 1959
Db 58538 GCGAGCGGTGCATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 58488
Qy |||::: |||
Db 1959 eLeu---AlaLysPro-----ProAl 1965
Qy |||::: |||
```

```
58487 ACTCAAAGCTCGGCTCGAGACCTTACCCCGGACTACACCCGCTACTTCCACCGGCTGG 58428
1965 aArgSerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuVa 1985
58427 CCGAACCGGTCC-GAGACCGGCTCTGTCGCGCCAGTGGCAGCTGCACNAGGCGATCGAC 58369
1985 lProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHi 2005
58368 GCGACACACCGCGCGCATCCACGACGAGTGTACCGGCGCACCTCGACGGCG----- 58314
2005 sHisAlaSerProAspProAlaProAlaSerAlaSerAspProHisArgGluLy 2025
58313 ----GCTGGCGCGACACCCCTCACCCCGCGCTCCACGTCGCGACCGCGGACGCGATC 58258
2025 sThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGl 2045
58257 GCCGCCACCGGTGCACTCCACC-----TGAGGAC 58225
2045 ySerSerTySerProGluGlyValGluProValSerProValSerProSerLeuTh 2065
58224 GAGCAGCAGAAGACCGCTCGCGCTCACACCGACGCGCTGCTCTCGCCACCGGCGAC 58165
2065 rHisAspLysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGlyGl 2085
58164 CGG-GAAGCGCGCTGGCGCGCTCTCTCGCGGCTCGACCCCTA-CATCGCGCGCGAC 58107
2085 uLeuArg-----ProLysGlnProGlyProValLysLeuGl 2097
58106 GCTCCGAACGCGCGCGTGCAGACCGCACCGCTCTGCTCTCGACCCCTCCGT----- 58052
2097 yGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSe 2117
58051 -----CACCGCTCGCTACTAGTCCAGAACGCGCGAGCAGCACACCC 58011
2117 rProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrLeuAlaGl 2137
58010 ACGGGTTCGCGACCGCGCGCTGGCGCTCGC-----CGCTGGCGCA 57969
2137 nHisIleSerGluValIleThrGlnAspTyrThrArgHisProGlnGln----- 2154
57968 G-----CGGACCATCTCTACGACCTCACCG 57942
2155 -----LeuSerAlaProLeuProAlaProLeuTySerPhe---ProGlyAlaSe 2170
57941 GCGGGAGCGCTACCGCTGCGCGCGCGACCGCTTACACACCTTGGCGCTGA----- 57887
2170 rCysProValLeuAspLeuArgPro---ProSerAspLeuTyThrLeuProProPro-- 2188
57886 -----CGGCGCTGCGCGCGAGATCCCGTTCGCGCGCGCGCACCGCGG 57846
2189 -----AspHisGlyAlaPro----- 2193
57845 CGCTACCCCGCTGTGTGAGCGGATCTAGGCGCGCTCGGGCACCGCTCAGAACACCGGC 57786
2194 -----AlaArgGlySerPro-----HisSerGluGlyGlyLysArgSerPr 2207
57785 ACGCGTGTGCGGTGTCAGCGACGAGTCTCTCGGTCTCGACGCGCGC 57726
2207 oGluProAsnLysThrSerValLeuGlyGlyGlyLysAspGlyIleGluProValSerPr 2227
57725 TTCGCGTCCCGCTCCGCGATCCGCGTCCGCGTCCGCGTCTCGCTCGACCTCGACACAGTCC 57666
2227 O-----ProGluGlyMetThrGluPr 2234
57665 TTGGCCGACCGCTCGCGGAGCGCGCGCGCTGGCGCGGTAGTTGTTTCACGCGC 57606
2234 O-----GlyHisSerArgSerAla----- 2240
57605 AGCAGCAACTGCTGGCGTGTCTCAGCGCGGCGCGCGGTAGTTCAGTTCGTCGT 57546
2241 -----ValTyrProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMe 2257
57545 GAACCGCGCGCTCGCGCA-----CGTCAATCTCTAGTACCGCGCGCGCACGCTAG 57495
```

Qy	2257	tGlySerLysSerProGlyAsnThrSer	-----GlnProProAlaAph	2271
Db	57494	TCGTAGTTGAGTCCGGGGCGCGTTCGGTGGTTCAGTTCACAGTCCACCGCGCA	57435	
Qy	2271	ePheSerLysLeuThrGluSer	-----AsnSer-AlaMetValLysS	2285
Db	57434	CCGGCGCGCGTGGCCAGCTAGTACTCCGCCAGTACTCCAGGTACGCCCGCAGCTGCCG	57375	
Qy	2285	erLysLysGlnGluLeuAsnLysLysLeuAsnThrHis	-----AsnArgAsnGluProG	2303
Db	57374	CCCGTCAGCAGCTTCGGCAGCAGCGTGTCTGTACAGTACAGGCTCCGACAGTCCCG	57315	
Qy	2303	luTyArgLysGlnProGlyThrGluLeuPheAsnMetProAlaIleThrGlyThrG	2323	
Db	57314	ATCGTCAGCTGC-CGGCCGGATCTCGAGGTCCGAGAGGGCGACCGCTGTCCGAG	57256	
Qy	2323	lyLeuMetThrTyArgSer-GlnAlaValGln	-----GluHisAlaSerThrAsn	2339
Db	57255	GACCGCAGCGAGCGGTACTCCGTGCCGCCAGCGCTCTTGTACACAGCTCCTCTGCAC	57196	
Qy	2340	MetGlyLeuGluAlaIleAlaIleArgLysAlaLeuMetGlyLysTyArgPheGlnTrpGluGlu	2359	
Db	57195	CTTGTGATCAGTGCATGATCGGGCGCTCTTGTACCGCGCTCCGACCG-TGCTCAGCG	57137	
Qy	2360	Ser	-----	2360
Db	57136	TCTCGTGGCGGTGCCACCACTGTTGACGTAGCGCAGCAGCTTCTCGTGTCTGTCG	57077	
Qy	2361	ProProLeuSerAlaAsnAlaPheAsnProLeu	-----	2371
Db	57076	CCAGCAGCTCGTGTATCCGGCGGTCTCTCGACCGGTGGAGTTGGCAGCGACGCGCT	57017	
Qy	2372	-----AsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGly	2388	
Db	57016	TACCGACTCGAGTGCAGCGCGCGCTCGAAGACCACTCGAAGTGAACAGCGGTGA	56957	
Qy	2389	-----ArgSerAspHisThrLeuThrSer	-----	2396
Db	56956	GCCGCTCGCGGAAGCAGCAGCGGTCCGACAGCAGCAGCTCTCCGGTCTCTCGTTGA	56897	
Qy	2397	---ProGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLys	2415	
Db	56896	CGACCTTCAGTTCGGCGATCTCTCGTGGCGGTGCGCGCAGCAGGATCGGTCTGATGCCG	56837	
Qy	2416	SerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSerValHisSer	2435	
Db	56836	GCACCTGCC-----GGGCCAGCTTGGCGCGCGGTCTCGACGTAGGGGA	56792	
Qy	2436	GluGlyAspCysAsnArgThrProLeuThr	-----AsnArgValTrpGlu	2451
Db	56791	CCTGTCACCGTAGGAGACGTGCCCGCAGCGCCGAGTCCCGCAGCAGCAGCAGCAGCT	56732	
Qy	2452	AspArgProSerSerAlaGlySerThrProPheProTyArgProLeuIleMetArgLeu	2471	
Db	56731	CCGCGCCCATCGAACGAGCTTCGGCACCCATT---TGCGCGCT-----	56690	
Qy	2472	GlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeu	2491	
Db	56689	-----GCTCTCCAGCGCCGGGAAGCGAGCTTCCCTCCAGCTAGC	56648	
Qy	2492	AlaGlyProHisHis	-----AlaTrpAspGluGluProLysPro	2504
Db	56647	CCTTGTCCAGATCGGATCGCCGGTGTGTTCAGCCCCCAGCAGCGCCA	56600	
RESULT	66			
LOCUS	AB089954/c	58343 bp	DNA	linear
DEFINITION	Micromonospora griseorubida gene cluster for the polyketide macrolide mycinamicin, partial and complete cds.			
ACCESSION	AB089954			
VERSION	AB089954.1	GI:28394148		
KEYWORDS				
SOURCE	Micromonospora griseorubida			
ORGANISM	Micromonospora griseorubida			
REFERENCE	1 Anzai, Y., Saito, N., Tanaka, M., Kinoshita, K., Koyama, Y. and Kato, F. Organization of the biosynthetic gene cluster for the polyketide macrolide mycinamicin in Micromonospora griseorubida FEMS Microbiol. Lett. 218 (1), 135-141 (2003)			
AUTHORS	Anzai, Y., Saito, N. and Kato, F.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-AUG-2002) Yojiro Anzai, Toho University, Faculty of Pharmaceutical Sciences; 2-2-1 Miyama, Funabashi, Chiba 274-8510, Japan (E-mail: yanzai@phar.toho-u.ac.jp, Tel: 81-47-472-2072, Fax: 81-47-472-2086)			
MEDLINE	22472038			
PUBMED	12583909			
REFERENCE	2 (bases 1 to 58343)			
AUTHORS	Anzai, Y., Saito, N. and Kato, F.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-AUG-2002) Yojiro Anzai, Toho University, Faculty of Pharmaceutical Sciences; 2-2-1 Miyama, Funabashi, Chiba 274-8510, Japan (E-mail: yanzai@phar.toho-u.ac.jp, Tel: 81-47-472-2072, Fax: 81-47-472-2086)			
FEATURES	Location/Qualifiers			
source	1. 58343			
	/organism="Micromonospora griseorubida"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:28040"			
gene	complement (1..910)			
	/gene="mycP"			
CDS	complement (<1..910)			
	/gene="mycP"			
	/codon_start=1			
	/transl_table=11			
	/product="methyltransferase"			
	/protein_id="BAC57022.1"			
	/db_xref="GI:28394149"			
	/translation="MSPSTGVLYLDDLLKRTVSNFYODATHVAGLITEAFVSEARESGEDYPTVAHTMIGMKRLNNLQHCVSALRDGDFDLETVGWGGACIFARGILKAYDVRDVTWVADSFGPKITDDDHDMDEAMNLHQVNEAVDLTSLATVQRFPSRYGLLDDQVFLFGWFKDTPATPERLAVLRMDGDSYGATMDVLTAVPRSPGGFAIDDDYCAPACRAVHEYDRHGSIDEIVELDRQGVYRSGLTFPVCRRSGRSRGTTCACHH TTGSSSGPRSNIPSLHRSVCVGPISRI"			
gene	1490..2641			
	/gene="mycCI"			
CDS	1490..2641			
	/gene="mycCI"			
	/codon_start=1			
	/transl_table=11			
	/product="cytochrome P450"			
	/protein_id="BAC57023.1"			
	/db_xref="GI:28394150"			
	/translation="MDRTCAWALPEQYAEFRQATCWPAKWWDGSPWLVSRVHVRA LVDPRVTPDTPQRLSEADGGDFRSMMLDPPETRLRMFISAPSVQVETMR PELEKVDGILDELLEALEPVDILTHALPMSTQVICHLLGVYEDREFFORSELAS RPNDKSPALLIELVEYLDGLVTKTAHPDTGLGTAVTERLLKGEITHQELVNNVAL LLAAGHETSANQVTLISVLTLRHPTAELRQPELMPNADVELLRHSHIADGLRRAA TADIVLGDHTIRAGDGLIILLSSANDHNTFGAETFDLHRPARHHVAFGYGPHQCLG QNLARLEMEVTGKLFRRVPAFLRLAQEPDALRVROGSPIFGIDELLVW"			
gene	2662..2886			
	/gene="mycCII"			
CDS	2662..2886			
	/gene="mycCII"			
	/codon_start=1			
	/transl_table=11			
	/product="ferredoxin"			
	/protein_id="BAC57024.1"			
	/db_xref="GI:28394151"			
	/translation="MRIVLDAERCVGAGCCEATAPELFTQDDGLGLVDRPVPVTELL GPAREAVDRCPVRAIRIESVVRGTWARGAG"			
gene	2883..3866			
	/gene="mydI"			
CDS	2883..3866			
	/gene="mydI"			
	/codon_start=1			
	/transl_table=11			
	/product="4-ketoreductase"			
	/protein_id="BAC57025.1"			

```

/db_xref="GI:28394152"
/translaton="MTRDRWNRSVVVTGALGFIGSHFVEBELAARGADVLGLRSEHR
AVDELSAHRVLLPVDLDERRLRQVDFHEASGAOTIVHCAALDGNAAVKAHSAE
ILDANLRAVNLNLCVRRDGVEDVTVLSSAAEYCGPTASPARDEDRCRTVRSGENGY
VLKLI TEILAEQVRRQYGGVHLVRPANVYGRDSDPGPASVIPAMIAESGGEI
EIWGDQQTSTFYVITDLVRAALALVETGKHSNLVNTTDETSMLDLARVVFSTGRT
ARIHKPAQPGAPGAVLDITTRKREVVDTYPTRTLEGLEETVRWYRHRIOR"

```

gene

/gene="mycE"

CDS

/gene="mycE"

/codon_start=1

/transl_table=11

/product="methyltransferase"

/protein_id="BAC57026.1"

/db_xref="GI:28394153"

```

/translaton="MTRQTEDEATVDVVLACGHDSERLTKQYDPAMI SRLIYA
EILSRCPSPNDTPVLVELAIVGSEFRHLVRVRSPIRPGADGFGVMLVEVEL
TELLRELFVTHERPAGVTKLFPYLTDDDEAVEQIGTYLLAAQDTEAVLACGCSR
KPLSELSSRYFTPKFGLHMTPTHYDRHFRYRQVRLVETGVGKYKHPWGGGSL
RMWKFPRQYIYGLDMDKSHVDELRIITIQDQNDABFLDRIARRYGPFDVIDDG
SHINAVRTSFAALFPHVRPGCLYVIEDMTAYMPGFGQADPOECSTGLLKLKSLI
DAIQHOELPSDPNRSPOYDNRNIVGLHVHYHNAFVEKGRNDEGGIPTWIPRDESLVQ
ASSGGAT"

```

gene

/gene="mycD"

CDS

/gene="mycD"

/codon_start=1

/transl_table=11

/product="glycosyltransferase"

/protein_id="BAC57027.1"

/db_xref="GI:28394154"

```

/translaton="MRTALITIGSRGDVQPIALGVGLRAGHDVILGAPGKHLVLE
RAGLITRPTGDPGPTMEVDLSRLSPYLRDLKALPAVEEDDQVVDIEAARG
QGADAVAPLTAAYGSDVDPWAAANWMTNTRFPAVSGQRRLGSLPLYLHF
THARAABEAWLREYDVRVRRQGLEPPFCGRSPFLRLGRERPLYLPFSFSLPKRD
WPARHVTYGFWDQWQPTRELVDLADGPAVADVRQHLAVHRRQVLEYIAIAA
RAAGRIIVVDGPDVPSDDVLHDVYRWLFPRTAAVHHGFGFTAEVVRAGRPH
VYVPFADHPFARRLTEIGVAAPAVDFTLTREALVASTVKAVREDSLGARAARLGE
AVRRGVENACQVSWESWIINSRTLTG"

```

gene

/gene="mycAI"

CDS

/gene="mycAI"

/notes="polyketide synthase"

/codon_start=1

/transl_table=11

/product="protomycinolide IV synthase 1"

/protein_id="BAC57028.1"

/db_xref="GI:28394155"

```

/translaton="MUVSVTAAPESVALVGMSCRLPGASGVBEFWDLITTEGRDAVAR
QPGSWRAALATGTFDAGFFGSPROAAAADPQORLVLLELHEALHAGVVPATLAG
SRIGFVGVGTADDTATSLMTGTTPDGTVPGLNRSMIANRLSLLELRGSPFVDSOA
QSSSLVAHLACSLLRGSDAIAVAGSLISTPESTAAWSLGLSPDRCYTFDAR
ANGYVRGEGAAVVKLSDALVDGRVHAVILGAVNNDGGPGGLIAPRAAQEEVL
RAYDRSGPAOVTVYELRGCTGVPVEAALGAVFCADRRADDPLHGVKSN
GHLEAGITGLLKAVLCVREGVLPPTLNHGTNGIPLAQLRLVQTVTVQRPDKN
EGRIIAVSFGMGCTNAHLVLQAQERAVRPVAVPTTELPAVPWLVSARSEALRGQ
AEKRLHLDRLDLRDLGFSLSATSRDVPFHRSVRPADGEPVAAAGSRLAERGVAWV
FPQGGQAWTMAASITESTESFPAARLTCFAAVTPFDVWSLDDVLRAVPGSPRLDRD
VVOQPAFCAVMVSLAEVWSGVPEAVVGHSGOEIAAAVAGAISELDGAIIVALKRSV
LTLRELHLGGMVSVGLPVAETERLLTRWSGRHLAAVNGPNATVAVSGADGALDELVT
ECERRQARVRVAEVAASHSPHYERIRDELMTALRPVTPQPSAVAFSTADYGTIGE
RLDAGTYRNLRQVPRADTLROMSTDGYTSFVELSAHFMVLNLEILAQETITALLC
VTGSLRRDEBGLGRMLASAAEAGHLFPVDMRKVFAGTGAERVELPTAPQRYWAP
DAERTVPPAADAVPAPPAASPEEQATPAQATTQOTVYDLRLTLETVAIVTQGLPSAE
VDPDLTFKEGGLSAAVEFRNLRAVGLSLPATVYDPTTVAAGFLGDQLHPRE
SOLVITPVALVDDEPTAIWGMGRFCFGVSSPEDLMDVVASGVDAVPPVPDVRWDLE
SLFDGPDPRAGTTCYREGGFLDDAAGDAGFFGVSPREAVAMPDQORLLLEGVGVG
AGGFSAGVRSGRTGVTAGVMNQYIGPRLYESAEHRGGLLTGTATAASVSGRVSYVIG
LEGPAVTVDPTACSSSLVLAHLVAQALRSCECDVALAGGVTVMATPGIFVFSRQGLA
VDGRCKAYSADSGTSGSEGVGVVVVQVRLSDAVAAGRVLAVVVRGSAVNDQGSNGIT

```

```

APSGSQQRVIRAAASARLSVGDVDVVEGHGTGTRLDPIEAQALLATYQRRGGGRSL
LLGSKVETNIGHAQAAGVAGVGMVAMERGVVPASLHVRCACFACQVSSGSAVEVVG
EACATPQRPRRAAVSSFGVGTNAHVIIEQAPVPVVDASAGSADVSSGADSSVS
VSPVPLSARSEWGLRAGACRLAALSGPLSGSGSGSVGVGVGWSLGRSVFGWR
GVSLSGVRLGVLAVGSGVGVVSGRVSASGVVFPVPCQAGWAGKRLGESGV
FSAPDEVGVQGLVDMVPVRGTACREVRGDLGRVDVQVPLSPVVMVALARLSRGV
VPPVGVHSGQGEIAAAVSGALVSEEGARVVGRSRVIGARLAGGAMVWVASGV
VEGLLGLGGLGGLVAVVNGPQVVVADGGGAGVGVGAGVGMGRARLLLPVGYGSHC
GLVDVREDLLGLVEVRGSSVPIFYSSVDVGVGDLGGVFRNLRLRVRFVGV
VRLVGDGSSFVEVSAHPVLGCAIEDVGEVGGVGVFCGVGVAVGVGGLGRMVVS
AAGSWVRGLPVQWRVPVLAGGRQVGLTYPPQHDRYWLAAPRTAIVAAGTADGAA
AGVGRTRVGMKALPPTPGLRHRMLVHGHATGGVQDLSGVVADILRTHGATTEHL
TIDPABPGVLAERLGAALAAAGEPVGVSLDLDGTLTLLVQAMGRAGLDARIVATTG
AVASDEVPSSEGAALWAFGRVAALELDRWAGLIDLPQADARALERLAGALTSEE
ENQAVRASGLYTRIRTAGRKKEWRPRGTVLVTGTGTGALGARVWRWLADHGAHL
VLVSRGAQATGAELVAGLERTGVTVVAAADVADRDRAGRAGGAPAGGATAGLPHWR
ARRDGPASHRSGSGGCPAERLAAHLDLTAHRLDADAFLVPSIVSGWNGAOAY
AANAADDALAEQRRAAGLPATSIAMGLWAGAGMADGAGQOTFTRRGVRAMPDGGIA
ALQALDSDGTCVTADVADWPMVRTHANPAARLFEEIPEARQALADRAAEQSAER
SDIARLRLTGLTAADRRRLLDLCAQAAVILRQDSQWTIIPDSAFRASGDFSLTALEL
RNRLNAATGLTLPSTVVFHPSVALDFLDLVLRSGGADGVVEVAAPADEPIAVI
GVSCRYPGVVATPEQLWDLLTERDAIGLPTDRGWNIDDI FDEPERGVRTYCREG
FLHDAADFDAAPFGISPREALAMPQORLILETSWEALERAGIDPRSLRSRTGYTG
MTHQEAARLHEMPEETEYEGHLLTGTSGVASGRVSYVIGLEGPVAVTDTACSSLVAI
HLAQALRAGECDLAGGVVMTATDCSSWSPASGSPRPVQVLRQADGTGWSG
AGVITLARISDAVRDGRHILALIRGPRSTVPATGSPRRTGRHSSASRLALANAGL
TPVEDVAEAGHTGRLGDI EAGALLATYQGRVPDQVPLRGLTKNIGTQAAAGV
AGMKVIALRHGVLPAHLHATSPSHIDMNAVASLLTERTDMPNPRRAAVSAF
GISGTNAHIIIEQAPETTPVPSERSAAGPRLSVVWPFLSGRSEPALRAQARLDRFL
DDGADPVVDGFSIATSRDLDRHRAVLLDGETAPTALHALVSGRSEPALRAQARLDRFL
ALAFLTGQSGQLGMDGIRREFPFPAEALAEIAAEMDTRLDPQLLSVMFAEPGSA
ATLIDRTYEAQALFAVEVALFRFPSPVPLVLLHSHSVEGIEIAAHVAGVMDIEHAC
RLVAARQLMORMPADGAMTARLPERVREQLDHLGRHVSVAAGVSVVVSAGH
VPLSADQLTAEGHRVRLNLSHAFSPHVDLEPERKAAAEVYGRPLVPVSVNT
GRVATEDLGSAEYWRHLRBPVRFDGINTLLDRGATTLVELGDPGLGCLAAQCLD
PDRDEVALPVALHRTSGVRSVLAATAVHVRGHDVDMNGFFAETGARRVDLPTAFQ
RQRYMASVPSGSDGTAAVRFGNLADHPLLGALSVAGSEIIFTGRLSLAQPMFA
GHVGRRAARAYAFLEVASRAGVGDVRDLQIEAPLVVPRGAVQVQVLTGDE
PDGQROVVARPEGSTPWRTHATGNGDAAGSGRRPGVGLAVPEGAPEGLDVLGR
PFAOHLVEDGVFTGVRELRWRADVEFVRLPDGAEEVAAAFGHVHALLDADLQPL
AGDFSVRSDSVHLFPANQMQVHTSGAALRVIRIRAGESA VSMQAVADGAPVCSL
DALVLRPTTVLRRTVPALLPLHLRMRERBPQAPGTLFTFLGTPGTVRVEYGEAV
TVHQDLTLARAVQGRGAPPRAVLAAGFSPHRTAEAVRAVAAALDLVQRWLADDRHQ
DSRLVLTERRAADDHEDVPDLGAAGVGLVRSQAQSEHPGFALIDHDGTEQSRALL
PAAVALGCAQLALRQGTVRVPAALAPPTESGADTRLAPDGTVLVTCGAALGALVVR
HLVAEHCVRLLILLSEHGEACWAELEADLQAGVTVTVLAQDADRCTGLAALADV
PDEPLITAVHAAGVLDMDMTGLTDRLDRDEVPRKVDAALHLHLTADRLKLAFLVF
SAAAGLIGNPGQASAAANTVLDALAHRRRANGLPATSLAWGLGGPNGAANLNAAH

```

Alignment Scores:

Pred. No.:	1,27e-05	Length:	58343
Score:	701.00	Matches:	596
Percent Similarity:	31.43%	Conservative:	287
Best Local Similarity:	21.22%	Mismatches:	1035
Query Match:	5.30%	Indels:	903
DB:	1	Gaps:	129

US-09-522-753-5 (1-2517) x AB089954 (1-58343)

Qy	16	GluProArgTyrProProHisSerLeuSer	-----TyrProValGln 29
Db	31881	CAACCGCGCCGCAACCCACCTCCACACCGCACCGCCCTCCACCCACCCAC	31822
Qy	30	IleAlaArgThrHisThrAspValGlyLeuLeuGluTyrGlnHis	-----44
Db	31821	ATCCCCCGCCACCCACCACTGCCCCCGGACCACTTAACCCGCCACCTTCCAAACACCCCCC	31762
Qy	45	-----HisSerArgAspTyrAlaSerHisLeuSer 54	
Db	31761	CAAAACCCCAACCAACCCCTCCACCAACCCACCGAGCGCCCGCCACCGACCATCGCCCC	31702
Qy	55	ProGlySerIleIleGlnProGlnA-ArgArgPro	-----SerLeu 68

Db 31701 ACCACCGCCACCGCGCCCTCACTACACGCGACCGACCCACCCACCGCGCACCCCTC 31642
 Qy 69 LeuSerGluPheGlnProGlyAsnGluArgSerGlnGluHisLeuArgProGluSer 88
 Db 31641 CTCACCGCACCGCGCCGACACCGCGCGGATCTCCCTCGGATGACCGAC 31582
 Qy 89 His-----SerTyrLeuProGluLeuGlyLysSerGluMetGluPheLeuSer 105
 Db 31581 CACGCGTGGGACACACACCGCCACGACCGCCACCGCGCCACCGCCACCCACCCAC 31522
 Qy 106 LysArgProArgLeuGluLeuLeuProAppProLeuLeuArgProSerProLeuLeuAla 125
 Db 31521 AAACGACACCGGCTGCACCATCAACCGCGCCCAATACCGCCACCGACCTCACCC- 31463
 Qy 126 ThrGlyGlnProAlaGlySerGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 145
 Db 31462 ACCGCGTGGCGGCA----- 31448
 Qy 146 GluProValSerProProSerPro-----ProHis-----ThrAspProGlu 159
 Db 31447 ---CGGCCAATCAACACCGCTGCACACCGCCACCGCCACCGCCATCATCAACGCGCTCGAA 31391
 Qy 160 LeuGluLeuValProProArgLeuSerLysGluLeuLeuLeuLeuLeuLeuLeuLeu 179
 Db 31390 ACACCTCGACTCCGCA----- 31373
 Qy 180 AspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLysGlnGln 199
 Db 31373 ----- 31373
 Qy 200 LeuGluGluGluAlaAlaLysProGluProGluProGluProValSerProProIle 219
 Db 31372 -----ACAAACCGCGAGCATCCCGCCACCTCCCGCGCGCTCGACCG 31331
 Qy 220 GluSer-LysHisArgSerLeuValGlnIleIleTyrAspLeuAsnArgLysLysAlaG 239
 Db 31330 GAAACACAAACAC-----CACACACCGCGACCGCCACCGCGCG 31292
 Qy 239 uAlaAlaHisArgLysLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnG 259
 Db 31291 AACCCACACCGCCAGAACCCACACCGCGCGCG-----CAACACACCGCA 31244
 Qy 259 nProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLy 279
 Db 31243 ACCCAACCGCCACCGACGACGACCGCGCCACCGCCACCGCCACCGCCACCGCCACG 31184
 Qy 279 sLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTyrLysPheCy 299
 Db 31183 A---CCACCCACATCCACCGCGGAAACAGACACCGACGACCGCGCGCGA----- 31132
 Qy 299 sGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAs 319
 Db 31131 -----AAGCGGACCGAAGCGCGCGCTCCA 31106
 Qy 319 nProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProG 339
 Db 31105 AACGACCG 31067
 Qy 339 uIleArgLysGlnArgGluGlnGluArgMetGlnSerArgValGlyGln-ArgGly 359
 Db 31066 CCGACACACCG 31007
 Qy 359 erGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleLeuAspGlyL 379
 Db 31006 CCG 30947
 Qy 379 euSerGlnGlnGlnAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetL 399
 Db 30946 TGCGGAAGAGGAAACCG 30903
 Qy 399 euTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAsp 419
 Db 30902 -----GCCACGCGACGCGCTCC 30884

Qy 419 roMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGlu 439
 Db 30883 CCAC-CACCTCCACCG 30834
 Qy 439 hrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuG 459
 Db 30833 CCCACATGCAACGAGCG 30798
 Qy 459 luArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsn 479
 Db 30797 ATCGCGCA-----CACCATCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30762
 Qy 479 yrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGln----- 495
 Db 30761 GCAGCAGCTGAGCATGACCAATATTCGACTTCACCGCGCGCGCGCGCGCGCGCGCG 30702
 Qy 496 -----GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 508
 Db 30701 CCACCGACCG 30642
 Qy 508 lngGlnProMet-ProArgSerSerGlnGlnGluLysAspGluLysGluLysGluLys 527
 Db 30641 GTCCCG 30582
 Qy 528 GluAlaGluLysGluGluGluProGluValGluAsnAspLysGluAspLeuLeuLys 547
 Db 30581 AACCGCGCGCGCG-----GluAspAsnAspGluLysGluAlaValAla 565
 Qy 548 GluLysThrAspAspThrSerGly-----GluAspAsnAspGluLysGluAlaValAla 565
 Db 30563 CGCTGTGCG 30504
 Qy 566 SerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgile----- 582
 Db 30503 GCGGAACCG 30450
 Qy 583 ThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAla 602
 Db 30449 AACCGCTGCG 30390
 Qy 603 GluLeuAlaSerMetGluLeuAsn---GluSerSerArgTyrThrGluGluMetGlu 621
 Db 30389 GAGAACGCTTACCG 30331
 Qy 622 ThrAlaLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetVal 641
 Db 30330 GACATCCGCGCGCGCG-----CATCACCGT----- 30307
 Qy 642 GlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsn 661
 Db 30306 -----CG 30262
 Qy 662 LeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArg 681
 Db 30261 CTGACCG 30202
 Qy 682 LysLys-----LysLysAla 686
 Db 30201 CGGCG 30142
 Qy 687 ProAlaAlaSerGluGluAlaAlaPheProValValGluAspGluGluMetGlu 706
 Db 30141 GTGCG 30082
 Qy 707 AlaSerGlyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAla 726
 Db 30081 CCCATTGGTGGCGCAT-----GTAGACCGCGCGCGCGCGCGCGCGCGCGCGCG 30041
 Qy 727 Ser-----GlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVal 742
 Db 30040 GCTCAACG 29996

Db	28196	TCGGCGATGTCCTCGGTCCGTAGC-----AGGAGGCGATCGGCG	28158
Qy	1366	laGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArgGluGlyThrProPro	1386
Db	28157	AGCAACAGGACCCAGCGTCCGTGAGGCGGC-----	28126
Qy	1386	roProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuL	1406
Db	28125	-----GFTCCGTTC	28116
Qy	1406	ysLeuLysProAlaHisGluGlyLeuValAlaThrVal-----	1418
Db	28115	ATCGGACCGGTGCCAACGTACGGGTACGGAGGTCTCTGCTGTGTCGGGGGTG	28056
Qy	1419	-----LysGluAlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrP	1437
Db	28055	GTCCCGCGGTGTGGGGGAGCCGACCTACCGGTCTGTGTAACGATACGTCCGCAAC	27996
Qy	1437	roGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProL	1457
Db	27995	CCACCTGACGCCACCCCAAGAACAGAACGCCCGCCGACCTGCACCCGGCAACCCCG-	27937
Qy	1457	eulysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeuI	1477
Db	27936	-----CACCAACCTCTGCGCGCATCTCCAC-----	27910
Qy	1477	leGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAla	1497
Db	27909	--CATCGGCCCAACACCATCTCCCGCGAGC-----	27880
Qy	1497	rgAlaLeuGluArgAlaCysTyrGluGlusLeuLysSerArgProGlyThrAlaSerS	1517
Db	27879	-----CAAGAACCCACACACAGCCACCCACAC-----	27850
Qy	1517	erSerGlyGlySerIleAlaArgGlyAlaProValIleValProGlu---LeuGlyLysP	1536
Db	27849	-----ACCCGCCACCTCCCGCCACACATCTCAATCGCC	27816
Qy	1536	ro-----ArgGlnSerProLeuThrTyrGluA	1545
Db	27815	CCCAACACCGGATGCGCGACACCTCCACAAACGACGAAACCCCATCACCACCAAC	27756
Qy	1545	spHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro---ValThrMetArg	1564
Db	27755	CG-----ACGCACACCTCTCCAAACGACCCCGCAACCCGCAATTCTTA	27711
Qy	1564	luProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAspArgL	1584
Db	27710	AACCAATACCCCATCCAAACCCACCATCCACACACCCACAT-CCACAGACGAATA	27652
Qy	1584	ysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHis-	1603
Db	27651	AA--ACGGAAATCACCGACTCAC-----CCGACAGAACTCTCCCGCAACACACT	27607
Qy	1604	-----HisProHisProIleSerProTyrGluHisLeuArgGlyValSerGlyV	1621
Db	27606	GAGCAACTATCTCCCGCACCATCCACCAACCCCAAT-----	27569
Qy	1621	alAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyI	1641
Db	27568	-----GCGACCATATACCCCGGCAACCAACCGCGCGCGCAACCCCA-----	27527
Qy	1641	leProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn-----P	1659
Db	27526	-----CTTCCCAACCGGCACCAACCAACCCCTC	27499
Qy	1659	roThrTyrProHisLeuTyrProPro-TyrLeuIleArgGlyTyrProAspThrAlaAla	1678
Db	27498	CACCCACCCCATCTCCCGCGCACCAACCATCTGCGCGG-----ACC	27457
Qy	1679	LeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsn	1698
Db	27456	ATTAACCCACCGCCACTCCAAACCAACCCCGCCCAACCCCGCCCAACCAACCCGCA	27397
Qy	1699	ThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSer	1718
Db	27396	CACCGACCGCCCGCACCATCTGCGCCCGCCACCGCCCA-ACCGCGCCCA-----	27344
Qy	1719	SerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValPro	1738
Db	27343	-----TCACCGCGACCGACCCACCA-----	27323
Qy	1739	HisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArgLeu	1758
Db	27322	-----CCACCGCGCACCTCTCCACCGCAACGCCC-----	27290
Qy	1759	AlaTyrLeuProThrAlaProGlnProPheSerArgHisSerSerSerProLeu---	1777
Db	27289	-----CCGACACCGCGCGCGGCACTCTCCCTCGCAATGACCCACCCCGCT	27239
Qy	1778	SerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGlu	1797
Db	27238	GCGGCACACACCCCGCAACGCCCAACCCCGCGCAACCGCCACCATCACCACAA	27185
Qy	1798	ArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleLeuThrSer	1817
Db	27184	-----ACG	27182
Qy	1818	ThrThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySer	1837
Db	27181	ACACCGCTGCACCAT-----CAACCGCGCCCAATGACCCCGCCCGACT	27134
Qy	1838	SerGlySerSerGlyGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAla	1857
Db	27133	CACCAACCGCTGCGCGCGCACCGGCAATCCCAACACCCCTGCACCCCGCCCAACCT	27074
Qy	1858	HisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSerVal	1877
Db	27073	CATCAACCGCTCCGAAACACTCTCGACTCCGCCA-----	27038
Qy	1878	LeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrVal	1897
Db	27037	-----ACAAACCGCGCGCATCTCCCGCCCGCCACTGCG	27008
Qy	1898	LeuArgSerThrSerThrSerSerProValArgProAlaAlaThrPheProAlaThr	1917
Db	27007	CCCTCTGACCGGAAACACAAACACACACACACCGCCGAC-GCCACACCGCCCGGACCCACC	26949
Qy	1918	HisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu	1937
Db	26948	ACCACCCCA-----GAACCCACACCA	26928
Qy	1938	LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThr---	1956
Db	26927	CCCGCGCAACACACCC-----AACCCACCCACACCCCGCAACCGACCCCGCCG	26877
Qy	1957	HisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSer	1976
Db	26876	CACCAAAACCGCAACGACCC-----AACGACACCCCGCATCCACACCCGAA	26826
Qy	1977	LysGlySerGluProArgProLeuValProValSerGly-----	1990
Db	26825	ACAGCACCAGAACCGAACC-----GAACCCGAAAGCGGACGAGAAAGCGCGCTCC	26772
Qy	1991	-----HisAla-ThrIleAlaArgThrProAla-----	1999
Db	26771	AAACGACCGCGCGCGCAAAACCCCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG	26712
Qy	2000	-----LysAsnLeuAlaProHisHisAlaSerProAspPro-----	2011
Db	26711	GCAACAGAGAGAACCGAGCTCCGAGACATCCGACACCGGACCCACCGCAGAGAA	26652
Qy	2012	-ProAlaProProAlaSerAlaSerAspProHisArgGlyLysThrGlnSerLysProPh	2031
Db	26651	TCGGCGCGCGCAACAGACACCGGAAACCAACAAAGAGCGGCGATCCGCAACCGGA	26592

Qy	2031	eSerIleGlnLeuLeuLeuArgSerLeuGlyTyrHis	2044
Db	26591	GCCTGCTCCAAATACACGTGGCGCTTCG-----TACCACCTAAACCCGGAACGACGACAC	26538
Qy	2045	-----GlySerSerTyrSerProGluGlyValGluProVa	2056
Db	26537	GCCCGCCGACCGCGCGACCGCTTCGCGCCAGCAGCACGCGCTCCCCACACCTCCAC	26478
Qy	2056	lSerProValSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluG1	2076
Db	26477	GCACCCGACGACCAATCCA-----CATCGAAGACGCGCACACCCACAT-----	26435
Qy	2076	uLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLe	2096
Db	26434	-----GCAACGAAGCGCGCACCA-----	26417
Qy	2096	uGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSe	2116
Db	26416	-----CCCCAC-----GCTCCATCGCCAAACACCATCTTC	26388
Qy	2116	rSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrLeuAl	2136
Db	26387	ACCAACC-----CAGCCACCCCGCAGCAGCCTGAGCATGACCAATATTCGACT-----	26339
Qy	2136	aGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSe	2156
Db	26338	-----TCACCGACCCCAACAACAAGACCGACCAACCAACGACGACCCATACGTC	26289
Qy	2156	rAlaProLeuProAlaLeuTyrSerPheProGlyAlaSer-----CysProVa	2173
Db	26288	GCCAAACACGCGCTCGCTCAATCGATCCCGCCAAACCGCGTCCCGTCCCATGCCCTCC	26229
Qy	2173	lLeuAspLeuArgProProSerAspLeuTyrLeuProProPro-----	2188
Db	26228	ACCACATCCACATCACCACCGCAACAAGCGCGACAGCGCAACGCGCGCGAATCACC	26169
Qy	2189	-----AspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysAr	2205
Db	26168	TGCTCGCAGCGACCACTCGGGCGCTCAAAACCATTCGACGACCATCTGATTAACGCC	26109
Qy	2205	gSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlylleGluProVa	2225
Db	26108	GAACCCCGCACCCCGCCAAACA-----CCCGACGACGACGACGACCC	26067
Qy	2225	lSerPro-----ProGluGlyMetThrGluProGlyHisSe	2237
Db	26066	GCATCCGACACCGCTGCACCAACACACACCCACCGCTCGGACCAACCGTACCATCA	26007
Qy	2237	rArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMe	2257
Db	26006	GCCGAATCCGAATACGCT-----TACACCGAC-----CATCCACGCGCAACCCACCC	25959
Qy	2257	tGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSer-LysLeuThrG	2277
Db	25958	TGCGCGCAAACTCCA-----CAAAAATCCCGCGCGTGGCATCACCGTCACCCCA	25908
Qy	2277	luserAsnSerAlaMetValLysSerLysGlnGluIleAsnLysLysLeuAsnThrH	2297
Db	25907	CCAGCCAAACCGCATCATCACTCCCGACAGCGCAACCGCTGCACCGCCAGATCAACGCC	25848
Qy	2297	isAsnArg-----	2299
Db	25847	ACCAACGACGAGAACACGCGTGTCCACCGTCAACCGCGCGCCCTCCAAACCAACACA	25788
Qy	2300	--AsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaI	2319
Db	25787	TAAGACACCCGACAGCAAAACGACACTGCGCGCGCTGCGGATCGCGAGAAAGCCCGCGTC	25728
Qy	2319	lThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrA	2339
Db	25727	TCCTCGTCGAAGGATGGTT-----CAGCGCGTAGTCGTGGTCATCAGCCCGCGCTAG	25674
Qy	2339	snMetGlyLeuGluAlaIleAlaLeuMetGlyLysAlaLeuMetGlyLysTyrAspGlnTrpGluG	2359
Db	25673	ACACCGGTACGGTCCCGCGCAACGC-----CAGCG-GATC	25639
Qy	2359	luserProProSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeu--P	2378
Db	25638	AAAACCCGCGCTCCAAACCGCTCCCGACACCTCCAAACCAACACGCTCTCGGGATC	25579
Qy	2378	roAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProG	2398
Db	25578	CATCGCCACACGCTCACGCGCGCACACCAAAAAACCCCGCATCG-----AAACCCG	25526
Qy	2398	lyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProA	2418
Db	25525	GCGCATCGACAGG-----AACCCACCGCTCCCGCAATACATCGCCCGCG	25481
Qy	2418	la---ProGlyLeuAlaSerGlyAspArgProProSerValSer-SerValHisSerGlu	2436
Db	25480	CACGCGCCGCG-----TCGGGATCGAAGATCGCGTCGACGTCCTCAGTGTGGTCCGTCG	25427
Qy	2437	GlyAsp-----CysAsnArgArgThrProLeuThrAsnArgValTrp	2450
Db	25426	GGAACGGGGAGATCGCGTCCGGAGGGGACAGCTCCCACTGTTCGGGCGCTTT	25367
Qy	2451	GluAspArgProSerSerAlaGlySerThrProPheProTyrAsn---ProLeuIleMet	2469
Db	25366	CCACCCACCCCGGAAC-----GGCAGCCCATTCACGATCGCCACCGCTCGTGTG	25313
Qy	2470	ArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGly	2489
Db	25312	CCTCTTCGGTCAACCGGTGCACCTGTGCCCTCGCTCGTGCAGGTCCGCCAGGACCGGC	25253
Qy	2490	ProLeuAlaGlyProHisHisAla-----Tip	2498
Db	25252	-----GAAGTACTACGCGAGCTTCTTCGTGTGGCAGTCACGCGCTTCCCTGG	25199
Qy	2499	AspGluGlu 2501	
Db	25198	GATCGTCAG 25190	
RESULT 67			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
AC026106			
LOCUS			
DEFINITION			
AC026106.12 GI:11094470			
VERSION			
KEYWORDS			
SOURCE			

[illegible]

Loulseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawlin, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Monabbat, K., Morgan, A., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okunnu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vazquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Unpublished
2 (bases 1 to 164485)
Worley K.C.
Direct Submission
Submitted (19-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 4, 2000 this sequence version replaced gi:9966554.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAM1
Center clone name: RP11-527C10
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-primer Bodipy: 9% of reads
Chemistry: Dye-terminator Big Dye: 91% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 146300 bases at least Q40
Consensus quality: 153355 bases at least Q30
Consensus quality: 156814 bases at least Q20
Estimated insert size: 154949; sum-of-contigs estimation
Quality coverage: 0% in Q20 bases; agarose-gel estimation
Quality coverage: 4.5% in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 36628: contig of 36628 bp in length
* 36729: gap of unknown length
* 68570: contig of 31842 bp in length
* 68571: gap of unknown length
* 93878: contig of 25208 bp in length
* 93879: gap of unknown length
* 93979: contig of 17701 bp in length
* 111680: gap of unknown length
* 111779: gap of unknown length
* 123986: contig of 12207 bp in length
* 123987: gap of unknown length
* 124087: contig of 9989 bp in length
* 134076: gap of unknown length
* 134076: contig of 7910 bp in length
* 142085: contig of unknown length
* 142086: gap of unknown length
* 142186: contig of 7295 bp in length
* 149481: gap of unknown length

Db 10212 ATCCACACCGCCGACCTCCAAACACCCACCGCCGCAACACCGGATGCCCGGAA-----CA 10156
Qy 89 sSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerLysArgProAr 109
Db 10155 CTCACAAACCCGAA----- 10140
Qy 109 gLeuGluLeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnPr 129
Db 10139 -----AACCATCCCCA-----ACAAAGCCGCC 10117
Qy 129 oAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluPro-Vals 149
Db 10116 CACCACATCGAAACAAACACCCCTCAGCAGATTCGGATACCAATACCCACCATCCAA 10057
Qy 149 erProSerProProHisThrAspProGluLeuGluLeuValProArgLeuSerL 169
Db 10056 ACCACCCCATCAACACACACACCCACCGAGCAATAAAACGACACACCCCGCCC 9997
Qy 169 ysGluGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGluGln 189
Db 9996 CGGAC----- 9992
Qy 189 lnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluAlaAlaLysProProG 209
Db 9991 -----ACAAACCCCAACACCCCGACAGCATCCCTCCACCGC 9952
Qy 209 luProGluLysPro-----ValSerProProIleGluSer-LysHisArgSer 225
Db 9951 CTCACATGCGCGGATCGAGCATATTCACCGCCACCGCGCGACCCACAC----- 9998
Qy 226 LeuValGlnIleTyrAspGluAsnArgLysLysAlaGluAlaHisArgIleLeu 245
Db 9898 ----- 9898
Qy 246 GluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGln 265
Db 9897 -----CCATCCCGCTCACATCCACCAACAAACCC----- 9868
Qy 266 TyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLys 285
Db 9867 -----CACAGCAGATCCATCCCGCGACACACCGTCACGCGGACCATTCAC-----CGC 9814
Qy 286 ArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeu 305
Db 9813 CGGCACACCAACCGACACCCACCGCCGCAACCGTTCCTGACATCCACGATCGGCAG 9754
Qy 306 MetGluAlaLeuGluLysLysVal-----GluArgIleGluAsnAsnProArgArg 323
Db 9753 TCGACCGACACCATCGACACACCCCGCCCAACACCTCACCGGACCGCGCA----- 9700
Qy 324 AlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGln 343
Db 9699 -----CCGCAACCGCCACCCCG 9682
Qy 344 ArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySer-----GlyLeuSer 362
Db 9681 CGCCCCATCGCACACTCAACACACACCCCGCCGCGAGCAACCTCACCTCGCA 9622
Qy 363 MetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGln 382
Db 9621 ATGACCCACACCGCATCCCGGACACACACCCCGCA-----CATCCACAAACCGCGCAA 9568
Qy 383 GluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAla 402
Db 9567 CGACCATCATCCGCCCGACAGACCGGCTGCACACATCAACCGGACCCACCCCTC 9508
Qy 403 AspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysVal 422
Db 9507 ACCCGCCCGCGCGCACCTCCAAAC-----ATCCCAATC 9469
Qy 423 TyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGlu 442
Db 9468 CACAAACCCCGA-----CAACGCCGACACACTCCACCAT 9433

Qy 443 LysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThr 462
Db 9432 CGACTCGCGCAAC-----CTCCGACTCGTCCAGCAGCGCGCAC 9391
Qy 463 ValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeu 482
Db 9390 CATACCCACCCACTG-----CCAACCTGACCGGAAACACAAACACCCACCCGAC 9337
Qy 483 Val-----ArgArgSerTyrArgArgGlyLysSerGlnGlnGln 496
Db 9336 ACACACGACCAACCCCGACCGCCAGAACCCACACACCGAACCCGAGCCACCGCCC----- 9278
Qy 497 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPro----- 513
Db 9277 CAGAAGCCACCGCAACCAAGACCAACCAAC-----TCCTCCACACCCCGCCACGACGAC 9219
Qy 514 ---ArgSerSerGlnGluLysAspGluLysGluLysGluLysGluAlaGluLysGlu 532
Db 9218 GCCGATGCTCAACACCGACCGCGACGACCAACCAACCAACCAACCAACCAACCA 9159
Qy 533 GluGluLysProGluValGluAsnAspLysGluAspLeuLysGluLysThrAspAsp 552
Db 9158 GAACACCCCGAG-----CCACACCCGTCAGAC 9129
Qy 553 ThrSerGlyGluAspAspGluLysGluAlaValAlaSerLysGlyArgLysThrAla 572
Db 9128 AACCCCGCGCACCTCACCAACCGACCGCGCATCGC-----CGCAAGCCCGC 9078
Qy 573 AsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
Db 9077 ACCGAACCGCGACACACCGCGCAACCGCAAGCTCTCTGCCAACACACCCCTCG 9018
Qy 593 GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer 612
Db 9017 CCGCCATCCAGACCGCGCATCGGAATCGGA----- 8985
Qy 613 SerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg 632
Db 8984 -----TCGGAATCGGAATCGGCA----- 8961
Qy 633 AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsn--- 651
Db 8960 -----GGATCGTCGAACCTCGGGGCGCTGCTCCAAATC 8928
Qy 652 -----PheTyrPheAsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnGln 668
Db 8927 AGATGGCATTCGTCC-----CGACACCCCAACCGCGACACACCGCGACGACGACGAC 8869
Qy 669 HisLysLeuLysMetGluLysGluArgAsn----- 678
Db 8868 CACACCGGCCAATCCCGCGCTCCCACTCCCACTCCACCCCGCGAGCAACCAATCCAC 8809
Qy 679 -----AlaArgArgLysLysLysAlaProAlaAlaAla 690
Db 8808 CAACGCGCGACGACATCCACATGCAACGTCCGCGCA-----CCACCCCGCGCGACGCA 8750
Qy 691 SerGluGluAlaAlaPheProValValGluAspGluMetGluAlaSerGlyVal 710
Db 8749 TCACCATTTGATCACACCGCGCACACCGCGCGCTGCTGATGACCAATATTCGACT 8690
Qy 711 SerGlyAsnGluGluMetValGluGluAlaLeuHisAlaSerGlyAsnGlu 730
Db 8689 TCACCGAACCAACACACCGCCCGCCACGACCGCCCGCCCGCCCGCCCGCCCGCC 8639
Qy 731 ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp-----Thr 748
Db 8638 -----CCAACACCGCTCAGCTCAATCGGATCAGCAAC 8603
Qy 749 GluSer-----IleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsn 765
Db 8602 GAGTCCCGCTCCATCGCGCTCCACCATTCACAT-----CCGACACACCCCAAC 8552

766	Qy	GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro	785
8551	Db	CCGCGCCCAACACGCGCAATCAACCGAGCTGCACCA-----CCCGCAGCGG	8498
786	Qy	ProArgArgThrSerArg-----AlaProIleGluPro	796
8497	Db	CCGTCAAAACATTTCGACCCACCATCTTGATTCAACGAGAACCCCGCACCA--CGGCCA	8441
797	Qy	ThrProAlaSerGluAlaThrGlyAlaProThrProProAlaProProSerProSer	816
8440	Db	ACACCCGATGACACACGCTCCCGCGCCACCGACCAACCGCTTCAACACGACACCCCGCAC	8381
817	Qy	AlaProPro---ProValValProLysGluGluLysGluGluThrAlaAlaPro	835
8380	Db	CCTCCCCCCCACCGCTCCCATCCGACACCCCGCAACGCGCCGACCGCCGATCACCCG	8321
836	Qy	ProValGluGluGlyGluGlu-----GlnLysProProAlaAlaGluGlu-----	850
8320	Db	ACAACCCACCTGCCGAGAAACTCCACAAACACCCCGCGCGCAATCACCGTCAACCC	8261
851	Qy	-----LeuAlaValAspThrGlyLysAlaGlu	859
8260	Db	CACCGCGCAACGCGCACATCACTCAACCCGCAACCCCTGACGCGCAATGCAACG	8201
860	Qy	GluProValLysSerGluCysThrGluGlu-----	869
8200	Db	CCACCAAGCAGCAGCAACACCGCGTATCAACCGNAATCGACGCGACCTCCAAACCCCAACA	8141
870	Qy	---AlaGluGluGlyProAlaLysGlyLysAsp-----AlaGluAlaAlaGlu	884
8140	Db	CATACGAAACACGACCGCGACGCTCAACCGACTTCGCGGTGAGCAGATACCCCTCGA	8081
885	Qy	AlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyLysSerGlyValGlyAlaThr	904
8080	Db	GTTCTTCGGAGCGCTCTGTCAGACGGGGCGCATCTCTGTTGGTTCAGCCCGGAGTACA	8021
905	Qy	ThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAsp	924
8020	Db	-----CGCGCGTCCGGCTACCCCGCAGAC-----CGCGCGGAT	7988
925	Qy	GluValAspGluAlaGluGlyLysAspLysAsnArgLeuLeuSerProArgProSerLeu	944
7987	Db	CGATTCCGCGCAGCT-----CCACGCTCCCGCAGC	7958
945	Qy	LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys	964
7957	Db	ACGTTCTCAACAACGAGCGTCTGCGGATCCATCGCACCGCTCACGCGCGGCGACACAC	7898
965	Qy	Gln-----LeuLysGlnArgAlaAlaAlaIleProProlleGlnVal	978
7897	Db	CAAGAACTCCGCATCGAACATCCCGCGTCAACAAACCCCGCTCCCGCGCATACG	7838
979	Qy	ThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPro-----	996
7837	Db	ACTTCCCGCGCGCCCGGACGGGATCAAGAGAGGAGTCCACGTCACCGCCCGCGTCT	7778
997	Qy	-----ProAlaProProProGlnAsn-----	1004
7777	Db	CGGGGCTGGGCGGATGACGTCAGCACCGCGCAGCAGGAGTCCGAGAGCTTCTCGGAG	7718
1005	Qy	-----LeuGlnProGluSerAspAlaPro-----	1012
7717	Db	TGTCGGCTCCGCGGGGTAGCGGAGCGCATGCCGATGGCAATAGGCTCTGTCGACGG	7658
1013	Qy	-----GlnGlnProGlySerProArgGlyLysSerArgSerProAlaPro	1028
7657	Db	GCACGCGTCCGCGCATCAACCGGA---CTTCTCTCGGAGAGCGCGCAGGAGCGCCT	7601
1029	Qy	ProAlaAspLysGluAlaPheAla-----AlaGluAlaGln	1040
7600	Db	CCAGGAACCTTCCCAACCTGGAGCGGGTGGTGGTTCGAAGACCAACCGTCGAGGACAC	7541
1041	Qy	Lys-----	1041

7540	DB	TCAGACCGGTCGTGCTGTTGAGTGGTTGCGCAGTTTCGAGCGCCGTGAGCGAGTCTGAAGC	7481
1042	QY	-----LeuProGly-----AspProProCysTyrThrSerGlyLeu	1053
7480	DB	CGCGCTCCTTGAAGGCCCGGTGGCGCGACCGCATCCACGCTCTCGTGCCTTCAGACCG	7421
1054	QY	ProPheProValProArgGluValLleLysAlaSerPro-HisAlaProAspProSe	1073
7420	DB	CGCGCGCGCCGCC-----GCACGTTTCGAGGAGGACCCCGCT	7382
1073	QY	rAlaPheSerTyrAlaProProGlyHisProLeu-----ProLeuG1	1087
7381	DB	GCCTGTCGCGTCGGGCATCCCGCCACCCGCTGCGCAAGCGCGGCTCGCGCATCGG	7322
1087	QY	yLeuHisAspThrAlaArgProValLeuProArgProProThrLleSerAsnPro-----	1105
7321	DB	TGTTCCAC-----TTTCGGGCATCAGCGCGCTCTGGCCTCGGCGACCCCGTCA	7271
1106	QY	----ProProLeuLleSerSerAlaLys-----	1111
7270	DB	AGAGCCGCCCGAGCGCATGTCGCAAGACGCGGAGCGAAACCTTCAGTCGACGTCGG	7211
1114	QY	-----HisProSerValLeuLgl-----	1119
7210	DB	CGACGGTGACAGACACGTCATCGCGTTCAGGCGCTGTCAGAGCGCGCATGCCGAGCT	7151
1120	QY	-ArgGlnLleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerG1	1139
7150	DB	CGGGTCCAAGGCCGTGATGCCACGGCGGTTTCAGGTTCTGCTCGCGCATCCCTCGCCA	7091
1139	QY	uHisAlaLysAlaProValGlyPro-----	1147
7090	DB	TGCCACCGCGGCCACATGTCGCCAAGCTATGAGGTGCGGGAAGACCTCGCGCACGAC	7031
1148	QY	----ValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerG1	1166
7030	DB	GTGCTCCGCCCGGGGT--CCAGCATGCTTCGCCCGCGCGTACGCGCCT-----GG	6978
1166	QY	yValLysGlnGluLnlLeu-----SerProArgGlyGlnAlaGlyProProGluse	1183
6977	DB	CCGCGTTGCCCGACACTCCGGTAACCGATGAGAAGAGACACGCGTCCAGCTCCTCG	6918
1183	QY	rLeuGly---ValProThrAla-GlnGluAlaSerValLeuArgGlyThrAlaLeuGlys	1202
6917	DB	TGCTGGGTGAGTTCGTGACACACACGCGGCACGCGTCTTCGACGGAAGACCTCGTCG	6858
1202	QY	erValProGlyGly-SerLleThrLysGlyIleProSerThrArgValPro-----	1218
6857	DB	ATGTGTCGCGCGGAGCGGTGTCGATC--ACACCGCTCGTGACGACTCCGCGGTGTGG	6801
1219	QY	-----SerAspSerAla-----IleThr	1224
6800	DB	AAGACCGCATGTGGCGGATCGCGCGCAGGACGTCGAGACGCGACGCCCGGTCCGTACA	6741
1225	QY	TyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThr	1244
6740	DB	TCGAGCGCGCCATGTCACCTCGACACCGAGC---CGCGAAGATCAGGTTCCAGTTCT	6684
1245	QY	ArgIleIleGlyGlnAspSerProSerArgLeuAspArgly-----	1258
6683	DB	GCCGCCCGGGCGTCCGCACCCGACGACTGCTGAGAACGAGGTGTTCCGACCGGTTG	6624
1259	QY	ArgGluAsp-SerLeuProLysGlyHisValLleTyrGluGlyLysGlyHisValle	1278
6623	DB	CGACGAGCACCCGGGCCACCCGCGCCACCCAGACTTCCCGTACCTCCGGTCACGAGATG	6564
1278	QY	uSerTyrGluGlyGlyMetSerValThrGlnCys-----SerLysGluAspG1	1294
6563	DB	TTCGCGGAGGGCGCCATTCGCGCGCGCATGCCCGCGCGCTCACCCGACGCCCG	6504
1294	QY	yArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMe	1314

Db 6503 TAGATCCCGGAGCGCGGTACCGGA-----TCTGGTCTCGCGTCGCGTCTC----- 6458
Qy 1314 tGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAl 1334
Db 6457 -----CGCGAGGGCGT-----TGCGGAGC 6438
Qy 1334 aileProProGluArgHisSerProHisHisLeuGluGlnHisIleArgGlyse 1354
Db 6437 TGGTCCAGCATCCGGCGATTCTG-----GTACGG----- 6410
Qy 1354 rIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgL 1374
Db 6409 -----ACGGAAGT----- 6401
Qy 1374 uAlaLysLeuLeuLysArgGluGlyThrProProPro-----ProPr 1389
Db 6400 -----CGACAGCGCGCCCGAGCGGTCCGGCAGTTCCAGGGCGGCGACC 6357
Qy 1389 oSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysPr 1409
Db 6356 CGCCCGAAGCGCCACACTCGCGCACCGAACACAT-----CC 6321
Qy 1409 oAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGlu----- 1427
Db 6320 GCGCCCTCTCTGGGGCACACACAGCGACCGCCCTCGGTGAGCTCCACAAAGGAGTCTCG 6261
Qy 1428 -IleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLys 1447
Db 6260 GCTCCCGCGTCACCGATGG-----CTGCACAGCGCCGAGCGAGACACGCCCATTTCCG 6204
Qy 1447 sGluGlySerIleThrGlnGlyThrProLeu-LysTyrAspThrGlyAlaSerThrThrG 1467
Db 6203 ATGGAGGGGTGTGAGGATGCGGTCTCTGCACAGCCGAG----- 6163
Qy 1467 lysrLysLysHisAspValArg-----SerLeuI 1477
Db 6162 -----AAGACACCGCGCCGACGAGCGGCACCGCCCGTCTGGTGGCGAGCAGT 6108
Qy 1477 leGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaA 1497
Db 6107 TCGCGATCCGGGCGCGCGCGGAGTTCACGCGCCGAGCAACTCGACCTCGGCGGCC 6048
Qy 1497 rgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGly----- 1513
Db 6047 CTGCGAATGAGCGTCTGT-----GATGCTGCGCAAGCTCCGAGCGCGGTCCCGCGG 5991
Qy 1513 ----- 1513
Db 5990 TCGCGGATGATAGCACGCGGTGGTGAGGGTCCGGGGTCTCGTCCGCAACGGTTTC 5931
Qy 1514 -----ThrAlaSerSerSerGlyGlySerIleAlaArgGly- 1525
Db 5930 CAGGTGATCCGATACCGACCGCCACCGCGGTCCGGGTGGGTGTCTCGCGGTGGCG 5871
Qy 1526 -----AlaProValIleValPro----- 1531
Db 5870 GGGTTGTGGCGGGTGGTGGAGTCCGTGCCAGTAGTGGCGGTGGAGACGGATAC 5811
Qy 1532 -----GluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProp 1550
Db 5810 GTCGGCAGGTCAACCAACCGCGCGCGCC-----CGCGAACACACACCTTCCAG 5760
Qy 1550 heAlaGlyHisLeuProArgLysSerProValThrMetArg-----GluProThrPro- 1567
Db 5759 TCAACTCAT-----CCGCGCAGAACGCGCTCCACCGACCGAGTCAAAAACCGCGCCAC 5703
Qy 1567 ----- 1567
Db 5702 CCACCATCATACGGCGAGGACCCACCACCGCGCGCATCCGGATCCACACC 5643
Qy 1568 -----ArgLeuGlnGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT 1586
Db 5642 GCCACGACTCCAAACAC-----CCACCGCCCAACACCGGATGCCCGGAACACTCC 5592

Qy 1586 hrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisPro- 1605
Db 5591 ACAAAACCCGNAACCCATCCCAACAAGCGCCCAACCATCGAAACAAACACACCCGC 5532
Qy 1606 -----HisProIleSerProTyrGluH 1613
Db 5531 TCACGCAGATTCCGATACCAATACCAACCAACCAACCAACCAACCAACCAACCAACCC 5472
Qy 1613 isLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheA 1633
Db 5471 ACCACCGACGA-----ATAAAACCGCACCAACCAACCAACCAACCAACCAACCC 5430
Qy 1633 spProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProA 1653
Db 5429 AAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 5370
Qy 1653 rgHisLeuAlaProAsnProThrTyrProHisLeuTyrPro----- 1666
Db 5369 GCATAATCCACCGCGCACCGCGCACCAACCAACCAACCAACCAACCAACCAACCAAC 5310
Qy 1667 -----ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgg 1683
Db 5309 CCCACAGCAGCATCCACATCCCGCGCACCAACCAACCAACCAACCAACCAACCAAC 5250
Qy 1683 lnThr-----llelleAsnAspTyrIleThrSerGlnG 1694
Db 5249 ACACCAACCGACCCACCGCGCAACGGTCTCTGGCATCCACATCGGAGTCCG 5190
Qy 1694 lnMetHisHisAsnThrAlaThr-AlaMetAlaGlnAlaAspMetLeuArgGlyLeu 1713
Db 5189 ACCGACACCATTCGCACCAACCGCGCAACCACTTACCAATACCGCGACCGCAACGCC 5130
Qy 1714 SerProArgGluSerSerLeuAlaLeuAsn-TyrAlaAlaGlyProArgGlylleleAs 1733
Db 5129 ACCACCGCGCGCCATCAGCCACACTCAACACACCCCGCCACCG----- 5084
Qy 1733 pLeuSerGlnValProHisLeuProValLeuValProProThr-----Pr 1748
Db 5083 -----CGCAGCAACTC-----CCTGGATGACCCACCGCATCCGCGACCAACGCC 5031
Qy 1748 oGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnPro-- 1767
Db 5030 ACCGACATCCACAACCGCG-----CCACGACACCATCACCGCC 4992
Qy 1768 -----PheSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLe 1785
Db 4991 CACGACACCGGCTGCACCATCAACCGCACCC-----ACACCCCTCACCAGCGCCACCC 4935
Qy 1785 uThrLysProThrThrThrSerSerSerGluArgGluArgAspArgGluArgAs 1805
Db 4934 CGCACACCTCCAAACAAATCCCAATCCACAA----- 4904
Qy 1805 pArgAspArgGluArgGluLysSerIleLeuThrSerThrThrValGluHisAlaPr 1825
Db 4903 -----ACCCGACAAACCGCGACCACTCC 4878
Qy 1825 oIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySerGlyGlyGly 1845
Db 4877 A-----CCATCGACTCAGAAAC-ACCTCGACTCTCCAGCAGCGCGCACCAT 4828
Qy 1845 yGlySerSerArgProAlaSerHisSerHisAlaHisGlnHis-SerProIleSerP 1865
Db 4827 ACCACCCACTGCACCCCTG-ACCCGGAACAAACAAAC-----CACCCGACCAACCCAC 4772
Qy 1865 roArgThrGlnAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysG 1885
Db 4771 CAGCAACACCGCGCGCACCGAACCC----- 4746
Qy 1885 lytleIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerS 1905
Db 4745 -----ACCACCGCGACCGCGAA-----GCCACCGCGCGCGGAACCAACCGCATCCA 4700

QY	573	AsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaIleSer	592
Db	14778	ACCGAAGCAGCCGACACACCGGAAACCGACAGCTCTGTCACACACACCCCTCG	14719
QY	593	GluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSer	612
Db	14718	CCGCCATCCACGACACCGGATCGGAATCGGAA-----	14686
QY	613	SerArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArg	632
Db	14685	-----TCGGAATCGGAATCGGAATCGGCA-----	14662
QY	633	AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsn----	651
Db	14661	-----GGATCGTCAACTCGGGGCGCTGCTCAAAATC	14629
QY	652	-----PheTyrPheAsnTyrLysLeuArgGluAsnLeuAspGluIleLeuGln	668
Db	14628	AGATGCGCATTCGTCC- CGACACCCCAACCGCGACACACCGCAGCAGCACGACC	14570
QY	669	HisLysLeuLysMetGluLysGluArgAsn-----	678
Db	14569	CACCACCGGCATCCCGCGCTCCGTCACTCCACCCACCCCGAAGACCAATCCAC	14510
QY	679	-----AlaArgArgLysLysLysAlaProAlaAlaAla	690
Db	14509	CAACGGCGAGCGGCATCCACATCAACAGTCCGCGCA-CCACCCCGCGCGACGCCA	14451
QY	691	SerGluGluAlaAlaPheProProValValGluAspGluMetGluAlaSerGlyVal	710
Db	14450	TCACCATCTTGATCACACCGCGCACACCGGCGCTCGGTATGACCAATATCGACT	14391
QY	711	SerGlyAsnGluGluMetValGluGluAlaLeuHisAlaSerGlyAsnGlu	730
Db	14390	TCACCGAACCACCAACACGCCCCCACCACGACCCCGCGCATACGACC-----	14340
QY	731	ValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp-----	748
Db	14339	-----CCAACAACGCTCAGCTCAATCGATCGATCAACCCCAAC	14304
QY	749	GluSer-----IleProSerProHisThrGluAlaLysAspThrGlyGlnAsn	765
Db	14303	GAGTCCCGTCCGATCGCCTCCACCATCCACAT-----CCGACACACCCCAAC	14253
QY	766	GlyProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProProThrPro	785
Db	14252	CGCGCGCACCAACCGCGCACCAATCACCGACGCTGCGCA-----CCCGGACGCG	14199
QY	786	ProArgArgThrSerArg-----AlaProIleGluPro	796
Db	14198	CCGTCAAAACCATTCACCCACCATCTGATTCACCGGAGAACCCCGCACCA-----	14142
QY	797	ThrProAlaSerGluAlaThrGlyAlaProThrProProAlaProProSerProSer	816
Db	14141	ACACCCGATGACCGCTCCCGCGCCACGACACCGCTCCCAACACACACCCCGCGCAC	14082
QY	817	AlaProPro-----ProValValProLysGluLysGluGluThrAlaAlaAlaPro	835
Db	14081	CCTCCCGCCACCGCTCCCATCCGACACCCCGCAACCGCGCACCGCCCATCAACCC	14022
QY	836	ProValGluGluGlyGluGlu-----GlnLysProProAlaAlaGluGlu-----	850
Db	14021	ACAACCCACCTGCGGAGAAACTCCAAACACCCCGCGCGCGCATCACCGTCAACC	13962
QY	851	-----LeuAlaValAspThrGlyLysAlaGlu	859
Db	13961	CACCGCCCAACGCACATCACACTCACCAACCGCAACCCCTGACACGCAAAATGCAACG	13902
QY	860	GluProValLysSerGluCysThrGluGlu-----	869
Db	13901	CCACCAACGACGACGACGCGGTATCAACCGAATCGACGACCCCTCCAAACCCCAACA	13842

QY	870	--AlaGluGluGlyProAlaLysGlyLysAsp-----AlaGluAlaAlaGlu	884
Db	13841	CATAGAAACAGCCCGACCGAGCTCACCGACTTCCGGTCGAGCAGATAGCCCTCGA	13782
QY	885	AlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyArgAlaThr	904
Db	13781	CTTCTCGGAGCCTCGTGCAGACGGCGGCATATCTCTGTTGGTTCAGGCGCGAGTACA	13722
QY	905	ThrAlaLysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAsp	924
Db	13721	-----CCCGCTCGGTACCCCGCAGC-----CCGCGCGAT	13689
QY	925	GluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeu	944
Db	13688	CGATTCCCGCAGCT-----CCAACGCTCCACG	13659
QY	945	LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys	964
Db	13658	ACGTCTCCAACAACGAGCGCTCTCGGATCATCGCACCGCTCAGCGGCGACACAC	13599
QY	965	Gln-----LeuLysGlnArgAlaAlaAlaIleProProIleGlnVal	978
Db	13598	CAAGAATCCGCATCGAACATCCCGGTATACACAAACCCCTCCCGCATACG	13539
QY	979	ThrLysValHisGluProProArgGluAspAlaAlaProThrLysProAlaPro-----	996
Db	13538	ACTTCCCGCGCGCGCGCAGCGATCAAGAAGAGGTCCACGTCACCGCCCGGTCT	13479
QY	997	-----ProAlaProProProProGlnAsn-----	1004
Db	13478	CGGGGCTGGGCGCATGACGTACAGCACCGGCCAGCAGGAGGTCCAGAGCTTCTCGGAG	13419
QY	1005	-----LeuGlnProGluSerAspAlaPro-----	1012
Db	13418	TGTGCGCTCCGCGGGTAGCGCGAGCCATGCCGACGATGCAATAGCTCTCGACGCG	13359
QY	1013	-----GlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPro	1028
Db	13358	GCACGCTCGCGCAGTCACCGGA-----CTTCTCGGACAGAGCGCGCAGGACGCT	13302
QY	1029	ProAlaAspLysGluAlaPheAla-----AlaGluAlaGln	1040
Db	13301	CCAGGAATTCGCAACCGTGGACCGGGTGGGTGGTTCGAGAACCCCTCGGAGGCGAG	13242
QY	1041	Lys-----	1041
Db	13241	TCAGACCGCTCGTCTGAGTGTTCGCGAGTTCGAGCGCGTGCAGAGTCGAAGC	13182
QY	1042	-----LeuProGly-----AspProProCysTrpThrSerGlyLeu	1053
Db	13181	CGCGCTCTTGAAGCGCGGGTGGCGCGACCGCATCCACCGCTCTGTCCTCAGGACCG	13122
QY	1054	ProPheProValProProArgGluValIleLysAlaSerPro-HisAlaProAspProse	1073
Db	13121	CCGCGCGCGCGCGC-----GCACCGTTTCGAGGAGGACCGCT	13083
QY	1073	rAlaPheSerTyrAlaProProGlyHisProLeu-----ProLeuGlu	1087
Db	13082	GCGTTCGCGCTCGGCGATCCCGCACCGCTCGCAGCGCGCGCTCGCGGACTCGG	13023
QY	1087	YLeuHisAspThrAlaArgProValLeuProArgProProThrIleSerAsnPro-----	1105
Db	13022	TGTCCAC-----TTTCCGGGCATCGAGCGGCTCTTGGCTCGGCGACCCCGCTCGA	12972
QY	1106	-----ProProLeuIleSerSerAlaLys-----	1113
Db	12971	AGACCCCGCGCGAGCTCGCAGAGCGCGAGCGGAAACCGTCCAGTCAGCTCGG	12912
QY	1114	-----HisProSerValLeuGlu-----	1119
Db	12911	CGACGCTGACAGACACGCTCATCGCGTCGAGGCGCTCTGTCAGAGCGCGCATGCCGAGCT	12852
QY	1120	-ArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlu	1139


```
Db 9260 -----
Qy 2362 oLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetPr 2382
Db 9255 AACCGGGCCCATCTCTGACCA---TCACCCCATATACACCGCTCCGGAACCC 9199
Qy 2382 oileThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLy 2402
Db 9198 CGCAAAACCAACGGAT---CAATCCCGCAGCTCCCAACGGCTCCC----- 9156
Qy 2402 sAlaLySValSerGlyArgProSerSerArgLyAlaLySValSerProAlaProGlyLeuAl 2422
Db 9155 -----ACGACACCTCCAAAACAAACCGCTGCTCGGATCCATCGCCACC-GCCTC 9107
Qy 2422 aserGlyAspArgPro----- 2427
Db 9106 ACGGGGGACACACCAAAACTCCGATCGAATATCCCGCGTCATACACAAACCCCCC 9047
Qy 2428 -----ProSerValSe 2431
Db 9046 CTCCCGCACATACGACTTCCCCACACACCCCGCAGCGATCAACAAACCCCTCCACATC 8987
Qy 2431 rSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGl 2451
Db 8986 CCAACCCCGATCCACCGGAACCCCGACACCGCATCCCGCCCAAC---ACCAAAATC 8930
Qy 2451 uAspArgProSerSerAlaGlySerThrProPheProTyr-AsnProLeuIleMetArgL 2471
Db 8929 CCACAAACCTTCCCGACACACCCACCCCGGAAAGACGACACCCCA----- 8883
Qy 2471 euGlnAlaGlyValMetAlaSerProProPro-----G 2483
Db 8882 -----TCCCCACACCAACACCGGATCATCCACGAGCAGCGGA 8843
Qy 2483 lyLeuProAlaGlySerGlyProLeuAlaGlyProHis----- 2495
Db 8842 GGTGACCTGCGGACGTTCTCCGGTTCGCGGCTGAAATTTCTTCTGCAGCTTCCCGGC 8783
Qy 2496 -----HisAlaTrpAsp 2499
Db 8782 AGCGGCCATCGGGTGGAT 8763

RESULT 70
LOCUS AC134881 204006 bp DNA linear HTG 01-OCT-2002
DEFINITION Homo sapiens chromosome UNK clone RP11-732A22, WORKING DRAFT
ACCESSION AC134881
VERSION AC134881.1 GI:23396432
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
WATERSTON R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 204006)
WATERSTON R.H.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H_NH0732A22

----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator; Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 198120 bases at least Q40
Consensus quality: 199344 bases at least Q30
Consensus quality: 200029 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1393: contig of 1393 bp in length
1493: gap of unknown length
2750: contig of 1257 bp in length
2850: gap of unknown length
4467: contig of 1617 bp in length
4567: gap of unknown length
5937: contig of 1370 bp in length
6037: gap of unknown length
7831: contig of 1794 bp in length
7931: gap of unknown length
10397: contig of 2466 bp in length
10497: gap of unknown length
13590: contig of 3093 bp in length
13591: gap of unknown length
15421: contig of 1731 bp in length
15521: gap of unknown length
19656: contig of 4135 bp in length
19756: gap of unknown length
26060: contig of 6304 bp in length
26160: gap of unknown length
31687: contig of 5527 bp in length
31787: gap of unknown length
36093: contig of 4306 bp in length
36193: gap of unknown length
46167: contig of 9974 bp in length
46168: gap of unknown length
58408: contig of 12141 bp in length
58508: gap of unknown length
129754: contig of 71246 bp in length
129854: gap of unknown length
202683: contig of 72829 bp in length
202783: gap of unknown length
202784: contig of 1044 bp in length
203828: gap of unknown length
203927: gap of unknown length
204006: contig of 79 bp in length.

FEATURES
Location/Qualifiers
1..204006
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-732A22"
1..1393
/note="assembly_name:Contig116"
1494..2750
/note="assembly_name:Contig119"
2851..4467
/note="assembly_name:Contig120"
4568..5937
/note="assembly_name:Contig122"
6038..7831
/note="assembly_name:Contig123"
7932..10397
/note="assembly_name:Contig124"

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
```


Qy	547	LysGluLysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSer	566
Db	146500	-----GCCAGC	146505
Qy	567	LysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMet	586
Db	146506	AAAGCTGGCCAAACAGCCAAAGCAGTCAGCAAGCCAGCCAGCCAGCTGGCCACCAGCCAA	146565
Qy	587	AlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSer	606
Db	146566	GCCAGACAAGCC---AGCCAAACAGCCAGACAGCCAGCCAGCCAGCCAGCCAGCCAGC	146622
Qy	607	MetGluLeuAsnGluSerSerArg---TrpThrGluGluGluMetGluThrAlaLysLys	625
Db	146623	CAAGAGACCCCAAGCAGCCAGCAGTGGCCAAACCCAGCCAGCCAGCTAGCCAAACCA	146682
Qy	626	GlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThr	645
Db	146683	GCC-----	146685
Qy	646	ValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIle	665
Db	146685	-----	146685
Qy	666	LeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLys	685
Db	146686	-----AAGCCGGCCACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCAGC-	146732
Qy	686	AlaProAlaAlaSerGluGluAlaAlaPheProValValGluAspGluLumet	705
Db	146733	AGCCCTGCCAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	146792
Qy	706	GluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHis	725
Db	146793	CAAGCCAGCCA-GCCAGCCAGCAGC-----CAAGCAGTCAGCCAGCCAGCCAA	146839
Qy	726	AlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSer	745
Db	146840	GCCAGC-----GAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA	146878
Qy	746	SerAspThrGluSerIlePro-----SerProHisThrGluAla	758
Db	146879	GCCAGCCAAAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	146938
Qy	759	AlaLysAspThrGlyGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGly---	777
Db	146939	AGCCAAAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	146998
Qy	778	-----ProProGlyProProThrProProArgArgThrSerArgAlaProIle	794
Db	146999	CAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147058
Qy	795	GluProThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProProSer	814
Db	147059	CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147115
Qy	815	ProSerAlaProProValValProLysGluGluGluGluGluThrAlaAlaAla	834
Db	147116	CAGTCAGCCAGCCAGC---CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147172
Qy	835	ProProValGluGluGlyGluGluGlnLysProProAlaAlaGluGluLeu-----	851
Db	147173	-----AAGCAAGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147214
Qy	852	---AlaValAspThrGlyLysAlaGluGluProValLys---SerGluCysThrGluGlu	870
Db	147215	CCAGCAACCCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147274
Qy	870	laGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGly	890
Db	147275	CGAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147334
Qy	890	laLeuLysAlaGluLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerG	910
Db	147335	ACAAACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147394
Qy	910	lyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaG	930
Db	147395	GCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147454
Qy	930	luGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGly	950
Db	147455	AG-----CCACTCAGCC	147466
Qy	950	spProArgAlaAsnAlaSerProGlnLysProLeu---AspLeuLysGlnLeuLysGln	969
Db	147467	AGCCAAAGATGCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147526
Qy	969	rgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArg---	986
Db	147527	CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147585
Qy	987	-----GluAspAlaAlaProThrLysProAlaProAlaProProProProG	1003
Db	147586	GCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	147645
Qy	1003	lnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLys	1023
Db	147646	AG-----CCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147681
Qy	1023	erArgSer-----ProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGln	1041
Db	147682	GCCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147734
Qy	1041	ysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProVal---ProProA	1060
Db	147735	-----CCAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147789
Qy	1060	rgGluValIleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProp	1080
Db	147790	AGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147834
Qy	1080	roGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProp	1100
Db	147835	CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	147894
Qy	1100	roThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGlu	1120
Db	147895	CA-----AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC	147948
Qy	1120	rg-----GlnIleGlyAlaIleSerGlnGlyMetS	1130
Db	147949	AGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	148008
Qy	1130	erValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrM	1150
Db	148009	GCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	148061
Qy	1150	etGlyLeuProLeu-----ProMetAspProLysLysLeuAla---ProPheSerGly	1167
Db	148062	--GGCCAGCTGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	148119
Qy	1167	alLysGlnGluGlnLeu-----	1172
Db	148120	CAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	148179
Qy	1173	--SerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGlu	1192
Db	148180	CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	148239
Qy	1192	laSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyI	1212
Db	148240	CCAGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC	148280
Qy	1212	leProSerThrArgValProProSerAspSerAlaIleThrTyrArgGlySerIleThrHis	1232

Db 4489 GCCTCCACCCGACTGGTCAGCGCCGACGACCATCCCGAG----- 4531
 Qy 2041 uGlyTyHisGlySerTySerProGluGlyValGluProValSerPro-----Va 2059
 Db 4532 -----TCAACCCCACTTGGTCCCGAGACCCGCTCCCTCCAGAGT 4575
 Qy 2059 lSerSerProSerLeuThrHis-----AspLysGlyLeuProLysH 2073
 Db 4576 GAACTCGCCCCCACTCCCATCCCAACCAACCACTTCCCGGACCTACCACTCCCGGCC 4635
 Qy 2073 isLeuGluGlu-----LeuAspLysSerHisLeuGluGlyG 2085
 Db 4636 ACCTTCGAGACCGTCTACCTCGAGTGGAGCGGAGCCCTCGCCCGCACCTTCTTGCC 4695
 Qy 2085 luLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuProH 2105
 Db 4696 GCCACCCACCGTTGGAGCTGGAGATCTT-----CTTCGC 4731
 Qy 2105 is-----LeuArgProLeuProGluSerGlnProSerSerProLeuLeuGlnThrA 2123
 Db 4732 AACAACTGGTCCAAACAGTTCCTTCATCAACCGCCCGACACCGTAGCGTCAACCT 4791
 Qy 2123 laProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValI 2143
 Db 4792 CT-----CACCTCGT----- 4802
 Qy 2143 leThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProL 2163
 Db 4803 -----CGCCCCCACCACCAACCAAAAGCAAGACCCACCTCTCCCA- 4844
 Qy 2163 euTySerPheProGlyAlaSerCysProValLeuAspLeuArgPro----- 2179
 Db 4845 -----CGTTCCCTCGCAAGCGTCTCGCTTCGCGACTCCACGAGCCCTACACCATC 4899
 Qy 2180 -----ProSerAspLeuTyLeuProPro-----ProAspHisGlyAlaProA 2194
 Db 4900 ACCGCCAAGGATCAAGCTTAGGTATCATCTCTTACCACCTCCCT-----CCAG 4947
 Qy 2194 laArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerV 2214
 Db 4948 CGCGCTTACTCGCTTCCCGGAGCGCTCCCTTCGCGACTCCGCTCTTCGCG-----G 5001
 Qy 2214 alLeuGlyGlyGlyGluAspGlyIleGluProValSerProGluGlyMetThrGluP 2234
 Db 5002 TCCTGTCATCGGAGAACGA-----CTTCGCCAGCTCACTCCAAACC 5046
 Qy 2234 roGlyHisSerArgSerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluP 2254
 Db 5047 CAAGCCACCATCCAAAGCAACGCTTCGCTC-----CGACCCGGATT-GGCGTCACAC 5099
 Qy 2254 roSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerL 2274
 Db 5100 CTTCGTCGATCTTCTCGAAGACCC-----AGCAAGGTCAACGAGAACTCTCTTTTCAC 5156
 Qy 2274 ysLeuThrGlu-----SerAsnSerAlaMetValLys-----SerLysL 2287
 Db 5157 CTCTTGAAGAGCTCCGACACCTCGCCCTCATCATGATTACTTGTCTCGTCTCTCGG 5216
 Qy 2287 ysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyAsnIleS 2307
 Db 5217 CCCAGTAAAAAAGATCAGCGCATCTCTCGACGCGGAGACCGCCCGCCCTCTCTCAT 5276
 Qy 2307 erGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrT 2327
 Db 5277 CCAGCGCGCGGACAGACCCCAACCACTTTCGAGTGTGTCGCAAAACCTCACTCCCTC 5336
 Qy 2327 yrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeu--GluAlaIleI 2346
 Db 5337 CGTCCACTCGCCATGACTACACGGCTTCGACCAATCGCAGCAGCGTAGCGGTGCTG 5396
 Qy 2346 eArgLysAlaLeuMetGlyLysTyAspGlnTrpGluGlu--SerProProLeuSerAlaA 2366

Db 5397 GCTCGAGGCT-----TGAAGATGCGCGCGCTTCATTCCTGA 5435
 Qy 2366 snAlaPheAsnPro-----LeuAsnAlaSerAlas 2376
 Db 5436 ACCTTTTCATCCCTCCACGTCACGTCGAAGACCAACATCGAATGCCAATTTGGCCTCT 5495
 Qy 2376 erLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrS 2396
 Db 5496 CAGCTGATCGCATCACCGGTGAACCGAGGACCTAGCAGCAACACCGACTCAACACT 5555
 Qy 2396 erProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysS 2416
 Db 5556 GGCCATCTCTACACCCAGTACTCTCCACCGCCCGCTCTGTGTCCGCGGAGA 5615
 Qy 2416 erProAlaPro-----GlyLeuAlaSerGlyAspArgP 2427
 Db 5616 CTCCCTCGTCAGCGCTCCCCCATGAACCTTCTTGGCTGCTCG-----C 5666
 Qy 2427 roProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThr- 2446
 Db 5667 CCGCTCT-----TCGCCCTCAAGCCCAACCCGAGACCTCCCGCTCGG 5711
 Qy 2447 -----AsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyRA 2465
 Db 5712 CTTTTCTGCGGTACTTCTCGCGCCCGCGCGCTCGCGCCCTTTT 5771
 Qy 2465 snProLeuIleMet-----ArgLeuGlnA 2473
 Db 5772 CGCCAGCTCGCATCGCTCGAGGACGCTCTCTGCCGAGAAATAGCTTTAGCT 5831
 Qy 2473 laGlyValMetAlaSerProProProGlyLeuProAlaGly-----SerGlyProLeuA 2492
 Db 5832 CGCTGAGTTCTCGTCTCGTCTCAGTCCCTCGCGACTCCTTTGTCCTCATCCCGCTG- 5890
 Qy 2492 laGlyProHisHisAlaTrpAspGluProLysProLeuLeuCysSerGlnTyGluT 2512
 Db 5891 -----AACTGTCATCTACCATGTCAGCTGCTTTGACCTCATCTCCGCCCA 5936
 Qy 2512 hr 2512
 Db 5937 CG 5938

RESULT 72
 AP000547 123288 bp DNA linear PRI 10-JUN-2000
 LOCUS Homo sapiens genomic DNA, chromosome 22q11.2, Cat Eye Syndrome
 DEFINITION region, clone:KB67B5.
 ACCESSION AP000547 BA000006
 VERSION AP000547.1 GI:5931525
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Shimizu.N.
 TITLE Homo sapiens chromosome 22 clone BAC KB67B5 on 22q11.2
 JOURNAL Published Only in Database (1999)
 REFERENCE
 AUTHORS Shimizu.N.
 TITLE Direct Submission
 JOURNAL Submitted (22-SEP-1999) Nobuyoshi Shimizu, Keio University, School
 of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
 160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp,
 tel:81-3-3351-2370, Fax:81-3-3351-2370)
 COMMENT This is a complete sequence of the insert of KB67B5 clone. The
 proximal adjacent clone is cl8E3 (Acc.#AP000546) with 4739-bp
 overlapping. The distal adjacent clone is KB7G2 (Acc.#AP000365)
 with 33343-bp overlapping.
 FEATURES
 source Location/Qualifiers
 1. .123288
 /organism="Homo sapiens"
 /mol_type="genomic DNA"


```

/evidence=not_experimental
/rpt family="AluSx"
30774. .30928
/evidence=not_experimental
/rpt family="Tigger3b"
complement(30931. .31193)
/evidence=not_experimental
/rpt family="AluSg"
31212. .31680
/evidence=not_experimental
/rpt family="Tigger3b"
complement(31713. .31878)
/evidence=not_experimental
/rpt family="MIR"
31987. .32277
/evidence=not_experimental
/rpt family="AluY"
32672. .32767
/evidence=not_experimental
/rpt family="n"(CGCGG)n"
33715. .33831
/evidence=not_experimental
/rpt family="L2"
34024. .34310
/evidence=not_experimental

```

Alignment Scores:

```

Pred. No.: 3.6e-05 Length: 123288
Score: 691.50 Matches: 602
Percent Similarity: 33.14% Conservative: 324
Best Local Similarity: 21.55% Mismatches: 1194
Query Match: 5.23% Indels: 679
DB: 9 Gaps: 123

```

US-09-522-753-5 (1-2517) x AP000547 (1-123288)

```

Qy 11 ThrTrpArgAlaThrGluProArgTyrProHisSerLeuSerTyrProValGlnIle 30
Db 73262 ACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGTCA 73321
Qy 31 AlaArgThrHisThrAspValGlyLeuLeuGluTyr-----Gln 43
Db 73322 AGCCAGCCATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCATCT 73381
Qy 44 HisHisSerArgAspTyrAlaSerHis--LeuSerProGlySerIleIleGlnGln 62
Db 73382 AGCCATCCAGCCATCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 73441
Qy 63 ArgArgArgProSerLeuLeuSerGluPheGlnProGlyAenGluArgSerGlnGluLeu 82
Db 73442 ACCGCCAGCCAGC-----CAGCCAGCCATGCCAGCCAGCCAGCCAGCCAGCC 73492
Qy 83 HisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPhe 102
Db 73493 ---GCCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 73546
Qy 103 IleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSer--- 121
Db 73547 ---CAAGCCAGCCAGCCAGCCAGCCAGC-----CAAAGCCAGCCATGCCAGCCAGCC 73597
Qy 122 -----ProLeuAlaThrGlyGlnProAlaGlySerGluAsp 134
Db 73598 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC 73657
Qy 135 LeuThrLysAspArgSerLeuThrGlyLysLeuGluProValSerProSerProPro 154
Db 73658 CAAGCCAGC-----CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 73693
Qy 155 HisThrAspProGluLeuLeuValProProArgLeuSerLysGluGluLeuIleGln 174
Db 73694 ---GCCAGCTAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 73732
Qy 175 AsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLys 194

```

```

Db 73733 -----:::|||||GCCAAGCTACCC 73744
Qy 195 LysLysGlnGlnGlnLeuGluGluAlaAlaLysProGluProGluLysProVal 214
Db 73745 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 73804
Qy 215 SerProProLleGluSerLysHisArgSer-----LeuValGlnIleIleTyrAsp 232
Db 73805 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 73864
Qy 233 GluAsnArgLysLysAlaGluAlaHisArgIle----- 244
Db 73865 AAAAATAA-AAAAAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 73923
Qy 245 -----LeuGluGlyLeu--GlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
Db 73924 AAAAATAGCTGGTGTGGTGGCAACACCTGTAATCCAGCCAGCTACTTGGGAGGCCCC--- 73980
Qy 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
Db 73981 -----AGGCAGGAGGATCGCTTGCATCCCGGAGGTAGAGGCTGCAGCCAG--- 74025
Qy 282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg 301
Db 74026 -----CCAGATTGCACCACTGCACCTCCAGCCTGG-----GCAACAGAG 74064
Qy 302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsn----- 318
Db 74065 TGAGACTCCATCTCAGAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 74124
Qy 318 ----- 318
Db 74125 CCCTGTCTCTACTAAATATACAAAAACATTAGCTGGACATGGTGGCACACGCTGTATA 74184
Qy 319 -----AsnProArgArg---ArgAlaLys 325
Db 74185 ACAGCTACTCGGAGGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 74244
Qy 326 GluSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGln-ArgG1 345
Db 74245 CGAACCAAGAAATTAGCTGGCGTGGTGCCACAGAAATAATATCCAGCTATTCGGGAGG 74304
Qy 345 uLeuGlnArgMetGlnSerArgValGlyGlnArgLysSerGlyLeuSerMetSerAl 365
Db 74305 AAAAAAAAAAAAAAAAAAGACAGCCAGCTGGCCACATGGTGAAACCGCGGCTCTACTAAA 74364
Qy 365 aAlaArgSerGlu-----HisGluValSerGluIleLeuAspGly-- 378
Db 74365 GATACAAAGAAATTAGCTGGCGTGGTGCCACAGAAATAATATCCAGCTATTCGGGAGG 74424
Qy 379 -----LeuSerGluGlnGluAsnLeuLysGlnMetArgGln-LeuAlav 394
Db 74425 CCAGGAGGAGGATCACTTTGAACCCCGGAGGTGGAGGCTGCAGCCAGCCAGCCAGCC 74484
Qy 394 allleProMetLeuTyrAspAlaAspGlnGlnArgIleLys----- 408
Db 74485 CACTGCACCCAGCCTGGTCAACAGGGCGAGACTGCATCTCAGAAAAAAGAAAAAAGC 74544
Qy 409 -----PheIleAsnMetAsn-----GlyLeuM 416
Db 74545 CTGGCCATCATGTGMAACCCCGCTCTACTAAATATACAAAAAATAGTTGGGGTGG 74604
Qy 416 etala-----Appro-----MetLysValTyrLysAspA 426
Db 74605 TGGCACACGCTTGTATCCAGGTACTCCAGAGCCAGCCAGCCAGCCAGCCAGCCAGCC 74664
Qy 426 rgGlnVal-----MetAsnMetTrpSerGluGlnGluL 437
Db 74665 GGGAGGTGGAGGTGGAGCTAAACAAGATTGCACCACTGCAGCTCCAGCTGGGCAACAGA 74724
Qy 437 ysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerP 457

```

Db 74725 GCAGACTCCCACTCAGAAAAAAGAAAAAGAAAAAACACCCAGCTG----- 74774
Qy 457 heLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLys----- 475
Db 74775 -----GCCAATGGTGAAAGACCGTCTCCACTAAATATACAAAAAATAGATA 74823
Qy 476 -----AsnGluAsnTyrLys----- 480
Db 74824 TGGGCATGGTGACACTTGTCTTAATCCCACTACTCGGAGGCCCATGTAGAGATCG 74883
Qy 481 -----SerLeuValArgArgSerTyrArgArgG 491
Db 74884 CTTGAACACAGAGGTGGAGCTGCAGCTAGCCAGTTTACACATTGCATCCAGCCTG 74943
Qy 491 lYlYsSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 510
Db 74944 GCAAAACCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75003
Qy 511 ProMetProArgSerSerGln-----GluGluLysAspGluLysGluLysGlu 528
Db 75004 CCAGCTGCGAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75063
Qy 529 AlaGluLysGluGluGluLysProGluValGluAsnAspLysGluAspLeuLysGlu 548
Db 75064 CCAGCCAGCCAGCCAGCCAGCCAG-----CTAAAGCCA 75096
Qy 549 LysThrAspAspThrSerGlyGluAspAsnAspGluLysGluAlaValAlaSerLysGly 568
Db 75097 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGTCA 75156
Qy 569 ArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgLysThrArgSerMetAlaAsn 588
Db 75157 GTGAAGCCAGCC-----AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75213
Qy 589 GluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGlu 608
Db 75214 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75273
Qy 609 LeuAsnGluSerSerArgTyrThrGluGluMetGluThrAlaLysLysGlyLeuLeu 628
Db 75274 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75324
Qy 629 GluHisGlyArgAsnTyrPheAsnTyrLysArgGlnAsnLeuAspGluLeuGlnGln 648
Db 75325 -----GCCAGCCAGCCAGCCAGCCAG----- 75345
Qy 649 CysLysAsnPheTyrPheAsnTyrLysArgGlnAsnLeuAspGluLeuGlnGln 668
Db 75346 GCCACCCAG-----CCAGCCATACAGCCCAAGC 75372
Qy 669 HisLysLysLysMetGluLysGluArgAsnAlaArgArgLysLysLysLysLysAlaProAla 688
Db 75373 CAC-----CCAAAGTCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 75420
Qy 689 AlaAlaSerGluGluAlaPheProValGluAspGluGluMetGluAlaSer 708
Db 75421 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 75480
Qy 709 GlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGly 728
Db 75481 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG----- 75525
Qy 729 AsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThr 748
Db 75526 -----CCAAACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA 75576
Qy 749 GluSerIleProSer-----ProHisThrGluAlaAlaLys-----AspThrGlyGln 764
Db 75577 GACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 75636
Qy 765 AsnGlyProLysPro-----ProAlaThrLeuGluValAspGlyPro 779
Db 75637 CAGGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 75696

Qy 780 ProGlyProProThrProArgThrSerArgAlaProIleGluProThrProAla 799
Db 75697 AAGCCACCAGCTAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 75756
Qy 800 SerGluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProPro 819
Db 75757 AAGACAGCC-----GGCCAGCCAAAGCGCGCAAG-CCACCCCGCCAGCCAGCCAGCCAGCC 75812
Qy 820 ProValValProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGlu 839
Db 75813 CAC-----CCGGCCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 75866
Qy 840 GlyGluGlnLysProProAla-----AlaGluGluLeuAlaVal 853
Db 75867 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 75926
Qy 854 AspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluGluGly 873
Db 75927 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 75980
Qy 874 ProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGluAlaLeuLysAla 893
Db 75981 CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76037
Qy 894 GluLysLysGluGlyGlySerGlyArgAlaThrAlaLysSerSerGlyAlaProGln 913
Db 76038 AGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76097
Qy 914 AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAsp 933
Db 76098 AAGCCACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76157
Qy 934 LysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAla 953
Db 76158 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76217
Qy 954 AsnAlaSer-----ProGlnLysProLeuAsp-LeuLysGlnLeuLysGlnAr 969
Db 76218 CAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76277
Qy 969 gAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAl 989
Db 76278 CAAAGTGGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76337
Qy 989 aAlaProThrLysProAlaProProAlaProProProGlnAsnLeuGlnProGluSe 1009
Db 76338 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76397
Qy 1009 rAspAlaProGlnPro-GlySerSerProArgGlyLysSerArgSerProAlaPro 1029
Db 76398 T-----CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76451
Qy 1029 roAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeu-----ProGlyAspProp 1047
Db 76452 CAAAGTCAACCAAGACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76511
Qy 1047 roCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerP 1067
Db 76512 CGAAGCAACCAAG-----CCAGCC-----AGCCAGC 76538
Qy 1067 roHisAlaProAsp-----ProSerAlaPheSerTyrAlaProProGlyHisProLeuP 1085
Db 76539 CAAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 76586
Qy 1085 roLeuGlyLeuHisAspThrAlaArgProValLeuPro-----ArgProProThrIleSerA 1104
Db 76587 ACTAAGCC-----AGCCAGCCAAACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAA 76643
Qy 1104 snProProProLeu-----IleSerSerAlaLysHisProSerValLeuGluArgGlnI 1122
Db 76644 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 76703


```
QY 34 HisThrAspValGlyLeuLeuGluTyrGlnHis-----SerArgAsp 48
Db 31350 ATGACGCGCGTGTCTCCAGCTCCACGACGACCAACGTCCTCCGCCAAGCGCCAC 31409
QY 49 TyrAla-----SerHisLeuSerProGlySerIleLeuGlnProGln--- 62
Db 31410 CGCGCCGCCAGGCGCGTGCAGACGACCTCGTGACACCAACGAGTGA-AGCGCCAGGTGA 31468
QY 63 -----ArgArgArgProSerLeuLeuSerGluPheGlnProGlyAsn 76
Db 31469 CCTGCTCGGAGCGCACCGCGCGGTCTTCGCAAGGTCCATGATC-----GGCGTC 31522
QY 77 GluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuPro---GluLeu 95
Db 31523 CCGCGGCGAGCGCGTCCGCGGCTCGGCCA-----GCCGTGGTCTCTCTCCGTGCA 31573
QY 96 GlyIysSerGluMetGluPheIleGluSerIysArgProArgLeuGluLeuLeuProAsp 115
Db 31574 GGAGCGCGCTCCACCGCGTCAGTTCCGCGACGCGTTCGCGGAGGTGCG-----CCTGCT 31627
QY 116 ProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeu 135
Db 31628 CCGCAATGCGGCG-----CCATCGCGGAACGAGAGCTCA-----31666
QY 136 ThrIysAspArgSerLeuThrGlyIysLeuGluProValSerProProSerProHis 155
Db 31667 -----CCGCGTCGAGCGAGCGCTCCCTCGCGCGCGCGCTCAC 31708
QY 156 ThrAspProGluLeuGluLeuVal-----ProProArgLeu 167
Db 31709 AGGACGCGCTCAGGTCCGCGAGCAGGATCTCCACGACCTCGGTCGACACCGCGGGGT 31768
QY 168 SerIysGluLeuIleGlnAsnMetAspArgValAspArgGluIleThrMetValGlu 187
Db 31769 GCGCCACCGACGAGCGCTCCGACACGCGGTCCCGGT-----31810
QY 188 GlnGlnIleSerIysLeuIysLysLysGlnGlnGlnLeuGluGluAlaAlaLysPro 207
Db 31811 -----CCA 31813
QY 208 ProGluProGluIysProValSerPro-----ProProIleGlu 220
Db 31814 CCCACACCGACGACCATCACGCGCGCGAGGATCCACGCGAGCGCGCTCCCGAG 31873
QY 221 -SerLysHisArgSerLeuValGlnIleIle-----TyrAspGluAs 234
Db 31874 CGCACCGCTCCGCGACGTCGTGAGGTACCGGTGTCGCTCGACCCGCTCGATCAGAT 31933
QY 234 nArg-----LysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnVa 252
Db 31934 GCGCGGGGTCCACCGCGCCCTCGCGGACAGCATCCCGGGGGTCCCGCGCCGACGG 31993
QY 252 lGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIl 272
Db 31994 TCGGCTCCGCAACATGCTCGTGGCGTCGATCATCCACGCTCAGCGCGCGCGCGCGTGT 32053
QY 272 eAsnGlnAlaMetArgIysLysLeuIleLeuTyrPheLysArg-----286
Db 32054 CCACCGACAGTCCGCGGCGCGCGAGCGTCCACCGCGCGGCGCGCGCGCGCGCGCGA 32113
QY 287 -ArgAsnHisAlaArgIysGlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMe 306
Db 32114 CCTTCATCGGCGTCCGCGGCGAGCTCTCCGNGTCTCGGGGGTTCGCGCGACCGCG 32173
QY 306 tGluAlaLeuGlu-----LysLysValGluAr 315
Db 32174 CGTCCAGGTCCAGATCCCGCGGATTCGCGCGGGTCTCTGTGTCGAACACCTCAGCG 32233
QY 315 gIleGluAsnAsn-ProArgArgArgAlaLysGluSerIysValArgGluTyrTyrGluL 335
Db 32234 CGCGAAGACCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 32293
```

```
QY 335 ysGlnPheProGluIleAArgLysGlnAArgGluLeuGlnGluArgMetGlnSerArg--- 353
Db 32294 CCCCG---CCCAACTCGAAGAAGTGTGTCATCCGCGCGCGACCCGATCCAGGTTCGAGAACCT 32350
QY 354 -----ValGlyGlnArg-----357
Db 32351 CGCGGAACAGCGGCACAGCGCGCTCGGCTGGTGACCGGTTCGCGCGCGGTGACGC 32410
QY 358 -----GlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValS 373
Db 32411 GTCGCGCGAAGTCCGGGTCCGCG-----AGCGCGCGCGGTCCACCTT-GCGGTG 32460
QY 373 erGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGlyLysGlnMetArgGlnLeuA 393
Db 32461 GCCGT-----CACCGGACGCGGTGAGGACGACGACCGCGCGC---32498
QY 393 laValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetA 413
Db 32499 -----CGGACCATGTAGTCGCGCGCGCGCGCTG 32526
QY 413 snGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpS 433
Db 32527 GCAGCGTCTCGCGACGCTTCGCGTC-----GACGTGCGCGCGCGACACACATAG 32580
QY 433 erGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyL 453
Db 32581 CCGATCAGGCGTCTTCACGTGCCACCGACCGCTGCTGTCGACGCGCGCTCAGCC 32640
QY 453 eulleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLeuT 473
Db 32641 AGCACGCGCTCGATCTCGCGCGTCCACACGCTAGCGCGGATCTTCACGTGTTGTCG 32700
QY 473 hrlYsLysAsnGluAAsnTyrLysSerLeuValArgSerTyrArgArgGlyLys 493
Db 32701 GCGCGCGCGCGTACACGCTCGCTCTCGTCC-AGTACGCGACGTGCGCGCTCG 32759
QY 493 erGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPro--- 511
Db 32760 GTACATGCGCTCACCGGAGCGAGCGGTCCGCGACGACCGTTCGCGCGCGCGG 32819
QY 512 -----MetProArgSerSerGlnGluGluLysAspGluLysGlu 524
Db 32820 ACGGTCCAGATAGCGTCCGCGCCACCGCGCGCGCGC-----GATGTAGAG 32864
QY 525 LysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAsp 544
Db 32865 ATCGCGCTCGACCGCGCGCAACGAGACGACATCCAGCAC-----32912
QY 545 LeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaVal 564
Db 32913 ---GTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 32969
QY 565 AlaSerLysGlyArgLysThrAlaAsnSerGlnGly-----ArgArgLysGlyArgIle 582
Db 32970 GCTGGGTTCGATCGCTTCAGGTTCGACACAGCGGTGCTCGGTGCGCGCGCGTGTG 33029
QY 583 -ThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSer-- 601
Db 33030 CCAGACGACGCTCGGGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 33089
QY 602 -----AlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGl 618
Db 33090 GTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 33148
QY 618 uGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerIleAl 638
Db 33149 GCCTCGGAACACTACCGCGCGTGAAGTT-----CACGGA-----GGTCAGCGCGCTCTG 33196
QY 638 aArgMetValGlySerLysThrValSerGlnCynLysAsnPheTyrPheAsnTyrLysLy 658
Db 33197 CGACATGAGCGCGCGCGCGCTC-----33218
QY 658 sArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAs 678
```

33219		33274	Db	34184	---CCACGAGCGTGTCCAGACCCCGCCCGCGAGCTTCTGGATCTCCGGGAGCCCGAGAT	34240				
678	n-----	AlaArgArg-----	Ly	682			QY	953	aAsnAlaSerProGlnLysProLeuAspLeuLysGln-----	Ar	969			
33275	CCAGAGTCTGTACAGGACCGCTCGAAGGTGTGCGGCGGTGCATCAGCAGCGTGTCTGC	33334	Db	34241	GCTGTGCGCGATCAGCGTGGACTGCGGTCTCTGAGCTCCGTGAGCGTCTCCATGACCG	34300	Db	969	gAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGluAspAl	989				
682	sLysLysLysAlaProAlaAlaAlaSerGluGlu-AlaAlaPheProProValValGluA	702	Db	34301	TTCTGCCCGCTCCAGCG-----	34348	Db	34301	TTCTGCCCGCTCCAGCG-----	CACGCGCACCGCAGCGGTGTGTGATGAACA	34348			
33335	CGCTCCGACGCGCCCGCGGTCTCCCGAGGCGCGACCGCGCGTGTGCGCAGCG	33394	QY	989	aAlaProThrLysProAlaProProAlaProProProGlnAsnLeuGlnProGluSe	1009	QY	989	aAlaProThrLysProAlaProProAlaProProProGlnAsnLeuGlnProGluSe	1009				
702	spGluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAag	722	Db	34349	-GCCCA-----	34401	Db	34349	-GCCCA-----	CCATGCCCTCCACGCCCGGAGTCCCGCGCGCGCGGAGACCGTCC	34401			
33395	CGACGCCCTTG-----	GGTGTCCCGTCTCCCGAGGTGTACATCAGTACGCCA	33445	QY	1009	rAspAlaProGlnProGlySerSerProArgGlyLysSerArgSerProAlaProPr	1029	QY	1009	rAspAlaProGlnProGlySerSerProArgGlyLysSerArgSerProAlaProPr	1029			
722	luAlaLeuHisAlaSerGly---AsnGluValProArgGlyGluCysSerGlyProAlar	741	Db	34402	GCGCGCAACATACGCTCTGCGCGCGCCAGTCTGGGCGCAGTCCGCGGCGCGGAGACCG	34461	Db	34402	GCGCGCAACATACGCTCTGCGCGCGCCAGTCTGGGCGCAGTCCGCGGCGCGGAGACCG	34461				
33446	GTGTCTCCGCCCGACCGGACGAGTCT-----	TCACCGTCGA	33487	QY	1029	oAlaAspLysGluAlaPheAlaGluAlaGln---LysLeuPro-----	1043	QY	1029	oAlaAspLysGluAlaPheAlaGluAlaGln---LysLeuPro-----	1043			
741	hrValAsnAsn-----	SerSerAspThrGluSerIleProSerP	754	Db	33488	CGGCATCGTCCATCGTGACGACACGTGGCGCGAGGCAACACGCGCGGTCCCGGCAC	33547	Db	34462	TGCACACCGTGTTCATCGTCAGCCCATGGCGCGGGGAACTCCGTTCAGGCGGAGGTG	34521			
754	roHisThrGluAlaAlaLysAspThrGly-----	GlnAsnGly-ProLysProProAla	771	QY	1044	-----	GlyAspProProCysTrpThrSerGlyLeuProPheProValPr	1058	QY	1044	-----	GlyAspProProCysTrpThrSerGlyLeuProPheProValPr	1058	
33548	CGCACACCGCGCGGACCGCGCGTCCGACATCATGAACGCGACCCGCTCCCGCGGT	33607	Db	34522	GAGGCGGTGAATGGCGATCACCTCCC-----	34569	Db	34522	GAGGCGGTGAATGGCGATCACCTCCC-----	GGACCGGAGTTCGCTGCC	34569			
772	ThrLeuGlyAlaAspGlyProProProGlyProProThr-----	ProPro	786	QY	1058	oProArgGluValIleLysAla---SerPro-----	HisAlaProAspProSe	1073	QY	1058	oProArgGluValIleLysAla---SerPro-----	HisAlaProAspProSe	1073	
33608	AGTCGGGTGCACGGGACGACGCGCTCCCGCTCCACACCGCGGACATCGCGCA	33667	Db	34570	CGCGCGCGCGTTCGAGATGTGAGCGCGATGATCGTCGGCACCGCCAGTTCGCG	34629	Db	34570	CGCGCGCGCGTTCGAGATGTGAGCGCGATGATCGTCGGCACCGCCAGTTCGCG	34629				
787	ArgArgThrSerArgAlaPro-----	IleGluProThr	797	QY	1073	rAlaPheSerTyraProProGly---HisProLeuProLeuGlyLeuHisAspThrAl	1092	QY	1073	rAlaPheSerTyraProProGly---HisProLeuProLeuGlyLeuHisAspThrAl	1092			
33668	CTACGTCGCGGAAGCTCCAGCAGGACACCGCGGTCCACACGTCGCGACGCCCGGT	33727	Db	34630	CGCCATG-----	34662	Db	34630	CGCCATG-----	CGCGCGCGCGCTCTTGTCT	34662			
798	ProAlaSerGluAlaThrGlyAlaProThrProProAlaPro-----	812	QY	1092	aArgProValLeuPro-----	ArgProProThrIleSerAsnProProPr	1107	QY	1092	aArgProValLeuPro-----	ArgProProThrIleSerAsnProProPr	1107		
33728	CCTTCAGTACCGCGCGCGCGCGCTCCGCTCTCGTCTCCGAGTCCGCGAACCAGCTTC	33787	Db	34663	CTCCCGACCATGCCAGATGTTCCGTAGGAGCGCGCGACCGCAGAC-----	34711	Db	34663	CTCCCGACCATGCCAGATGTTCCGTAGGAGCGCGCGACCGCAGAC-----	34711				
813	-----	ProSerProSerAla-----	817	QY	1107	oLeuIleSerSerAlaLysHisProSerValLeuGluArg-GlnIleGlyAlaIleSerG	1127	QY	1107	oLeuIleSerSerAlaLysHisProSerValLeuGluArg-GlnIleGlyAlaIleSerG	1127			
33788	GTCCCCCGCGTCATCGCGCTCGCGTCCGGGCGCTCTCCACCTGCGCGGAGAGAT	33847	Db	34712	-----	CCGACAGCATTCGCGCGCGCGGTAGATCTCGAGGAGT	34749	Db	34712	-----	CCGACAGCATTCGCGCGCGCGGTAGATCTCGAGGAGT	34749		
818	ProProProValValProLysGluGluGluGluThrAlaAlaAlaProProVal	837	QY	1127	lnGlyMetSerValGlnLeuHisValProTySerGluHisAlaLysAlaProValGlyP	1147	QY	1127	lnGlyMetSerValGlnLeuHisValProTySerGluHisAlaLysAlaProValGlyP	1147				
33848	CCACCGCGCGCGCGCGTCCGCTCTCGTCTCCACCGTTCGACACCGACCGCTC	33907	Db	34750	TCGGAATCGAGATCCCGCGCACCGCTCCATGAGCAGGTGGTGGCTCGTGTACAG	34809	Db	34750	TCGGAATCGAGATCCCGCGCACCGCTCCATGAGCAGGTGGTGGCTCGTGTACAG	34809				
838	Glu-----	GluGlyGluGluGlnLysProProAlaAlaGluGlu	850	QY	1147	roValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyV	1167	QY	1147	roValThrMetGlyLeuProLeuProMetAspProLysLysLeuAlaProPheSerGlyV	1167			
33908	CGCGTTCGCGTCCGCGCAGGAGTTCAGCGCGCGCGCGCGCTCCGCGTCA	33967	Db	34810	AGACA-----	GTGCGCTTCGCGCGCGGATGAGCGTCAGCCG-----	34850	Db	34810	AGACA-----	GTGCGCTTCGCGCGCGGATGAGCGTCAGCCG-----	34850		
851	LeuAlaValAspThrGlyLysAla-----	GluGluPro-----	861	QY	1167	alLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeu-----	1184	QY	1167	alLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeu-----	1184			
33968	TCTGTCTCAGGACGCGGACGAGCGCGTGGAGCGCGCTCCGCGACTCCGCGCTCAG	34027	Db	34851	-----	CAGCAGCGGAGCGCGCT---CAGTCCGAACCCCTCGCGCTCTTCTG	34899	Db	34851	-----	CAGCAGCGGAGCGCGCT---CAGTCCGAACCCCTCGCGCTCTTCTG	34899		
862	-----	ValLysSerGluCysThrGluGluAla-----	GluGluGly	873	QY	1185	-----	GlyValProThrAlaGlnGluAlaSerVal-----	1194	QY	1185	-----	GlyValProThrAlaGlnGluAlaSerVal-----	1194
34028	TCTCCGCTGTGTGATCTCGATCTCTCAGCGCGATGAGCGCGAGTGA	34087	Db	34900	GCGATCAGCTGACCTCGGCCATCGCGTCTCTCGCTCAGCGCGGACAGTCCGCTCT	34959	Db	34900	GCGATCAGCTGACCTCGGCCATCGCGTCTCTCGCTCAGCGCGGACAGTCCGCTCT	34959				
874	ProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAla	893	QY	1195	-----	LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleT	1209	QY	1195	-----	LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleT	1209		
34088	TCGCGTAGGGG-----	ACGCTGTGTGAACCTCT-----	34117	Db	34960	GTCGAGGAAGTTGGACATCCCGACATGAGTGGACCGTCTCGCGGACTTCCGTCGG	35019	Db	34960	GTCGAGGAAGTTGGACATCCCGACATGAGTGGACCGTCTCGCGGACTTCCGTCGG	35019			
894	GluLysLysGluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGln	913	QY	1209	hrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyArg-----	1226	QY	1209	hrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyArg-----	1226				
34118	-----	GAGCGCGCGGGGACTTCGAGCCCGCGCG	34162	Db	35020	TGGAAGGACCGCGGAGCGTGGGTGCGGTGCGGTGCGAGGAGGCGCTGCCAAGACGTC	35079	Db	35020	TGGAAGGACCGCGGAGCGTGGGTGCGGTGCGGTGCGAGGAGGCGCTGCCAAGACGTC	35079			
914	AspSerAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyAsp	933	QY	1226	-----	-----	1226	QY	1226	-----	-----	1226		
34163	GATCGACGTAGTCTCGAACA-----	-----	34183	Db	35080	CGCTCGGTGTCCAGCGCGCTGTGATGATCTGGGTCTGTGGATGACGTAGATGTCGGGG	35139	Db	35080	CGCTCGGTGTCCAGCGCGCTGTGATGATCTGGGTCTGTGGATGACGTAGATGTCGGGG	35139			
934	LysAsnArgLeu-LeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAl	953	QY	1227	-----	GlySerIleThrHis-Gly-----	1232	QY	1227	-----	GlySerIleThrHis-Gly-----	1232		
				Db	35140	GCCGCGTCTCGAAGGAGGAGTGTGATAGCATTCCTCTCTGTCATCGGGGACAGCGCCAG	35199	Db	35140	GCCGCGTCTCGAAGGAGGAGTGTGATAGCATTCCTCTCTGTCATCGGGGACAGCGCCAG	35199			

```
Qy 1233 -----ThrProAlaAspValLeuTyrLysGlyThrIleThrArgIle 1247
Db 35200 ATCTCGGCGAGCGCGATCCAGGAGAGTGTGCTGGGTACGGTACCGACCTC--- 35256
Qy 1248 GlyGluAspSerPro-----SerArgLeuAspArgGlyArg 1259
Db 35257 ---CGCGAAGTCGTCGCTATGGCTTTCAAACCCCTCGATCTCGTCTCGATCGAGGTCGA 35313
Qy 1260 GluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGlyHisValLeuSer 1279
Db 35314 GAA----- 35316
Qy 1280 TyrGluGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGly 1299
Db 35317 -----GGTCGAAGTCGAGCGGTGTGCCCGCGCGAGGAGTGTCTCCGCCAGGT 35367
Qy 1300 ProProHisGluThrAlaAlaPro-----LysArgThrTyrAspMetMetGluGly 1316
Db 35368 GGGCCAGCGCGGACAGCGGTGTCCGCCACATCCGACCGAACCGTTTCGATCTCGGCCCTCTT 35427
Qy 1317 ArgValGlyArgAlaIleSerSer-----AlaSerIleGluGly-LeuMetGlyArgAl 1334
Db 35428 CGAGGAGCGCGCTGCCAGCTCCCATGAGCTTCAGTTTCGGGCCCTG-----CGCA 35481
Qy 1334 alleProProGluArgHisSerProHisHisLeuLysGluGlnHis---HisIleArgG1 1353
Db 35482 TGTCCCGGACGACCACTGCAGCTGCAGCGATGCGCGCAGTGCATGCGCGGTTCAGGG 35541
Qy 1353 YSer-----lleThrGlnGly11 1359
Db 35542 AACCGCGGATGTACCGCGCTGTGCCACTCCGCGCAGTGTCTGTCTGCCCGCCAGCGCA 35601
Qy 1359 eProArgSerTyrValGluAlaGln-----GluAspTyrLeuAr 1372
Db 35602 ACCGGCCCATGTAGTTGAATCCGATCCAGGGGACGCGATGGCTCCAGCAGCGGCCCG 35661
Qy 1372 gArgGluAlaLysLeuLysArgGlyGlyThr-ProProProProProSerArgA 1392
Db 35662 TCGCGGGTTCAGATGCGCAACAGCGCGTGACCGAGCCGTGCATCAGCAGCCCGCGTG 35721
Qy 1392 spLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisG 1412
Db 35722 ACTGCTCTTACCGCGCTTACACAGATCC-----CCGACGCCA 35760
Qy 1412 LuGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluG 1432
Db 35761 GGTGATTCCTCCGAGACGTCCAGGGCGCACCGGGTGACGCTGGTGAACACGACCGCGTCC 35820
Qy 1432 LuLeuArgHisThrProGluLeuProLeuAlaProArgPro---LeuLysGluGlySerI 1451
Db 35821 GGGACAGATCAGCCCG-----CCGACCGGTGACCGCGGTGACTCTCCACGTCCACCA 35874
Qy 1451 leThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysH 1471
Db 35875 GCACCAACACCGCGCGCGGAGCAGCGCGCCACCGCGCGCGCAACCCCGCGAGC----- 35927
Qy 1471 iAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHiPro--- 1489
Db 35928 -----AGTATCTGTGACCCCGCAGAGAACACGCCCGCGGAACAGTCTCCACGA 35976
Qy 1490 --LeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuL 1509
Db 35977 GTCCGACGCTCGGCTCCCGCATATGACCCAGCAGCTGTAGAGCCACACACTCCCGGTGCTCC 36036
Qy 1509 ySerArgProGlyThrAlaSerSer-----SerGlyGlys 1521
Db 36037 CTGCGCGCGCTGCACCGCGGTGTCCACCTCGGTTCAGGATCGCTGCGCGCGCGAGTT 36096
Qy 1521 erIleAlaArgGlyAlaProValIle-ValProGluLeuGlyLysProArgGlnSerPro 1540
Db 36097 CGCGACCGCGCTCGCGGAGTGTGCCCACTCCACGACGAGGTGTGCCCGCGCGGAACG 36156
Qy 1541 LeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
Db 36157 -----GCACCGCCACCGGCTCCAGCGCAGGGGTCCGTCCCGCAGGACCGCTCCG 36207
Qy 1561 Thr---MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerLysAla 1579
Db 36208 ACCCGCGCGCGATCCGGCACCA-----GAATCCGCAGGACA 36246
Qy 1580 SerGluAspArgLysLeuThrSerThrProArgGluIleAla----- 1593
Db 36247 CGCGTTCGACGGCATGTGTGCACCGCATACCAATCGCCCGCCACCGTTCGCGCCCG 36306
Qy 1594 -----LysSerProHisSerThrValProGluHis 1603
Db 36307 CGTCCACCCACACCGCTGCACCATCACGCCGTCTCCGGTCCAGCGCTCCACCGCGT 36366
Qy 1604 ---HisProHisProIleSerProTyrGluHisLeuLeuArgGlyVal-----SerGly 1620
Db 36367 CCGCACCGCGGCTCCAGCGCTGTCCAGCAGCGCTCTCGTCGAGCCCGCCACCGCTCCGGA 36426
Qy 1621 ValAsp-----LeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
Db 36427 CCGACCGCGCGTGTCCAGGAGCCAGTTCACCTAGA-----TCAGGCGTCCC 36477
Qy 1639 ArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn 1658
Db 36478 ---GCTCCACCGCTCGCAACATGTGTGGTGTGCACCGCGCGCCAGCGCGCG 36534
Qy 1659 ProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAla 1678
Db 36535 CCAGGATCCCTCAC-----CCAGATCCCGCGGTGCGCCCAACACCCACTGCCGA 36588
Qy 1679 LeuGluAsnArg-GlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAs 1698
Db 36589 ATCCCGACCGGTGATCCGTC-----GCCAGGTTTCGATCACCG 36630
Qy 1698 nThrAlaThrAla---MetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgG1 1717
Db 36631 -GGTCCACGGCACCTCGCGCGTACCGTCCGCGCGTCCGCGCGCGAGCGTTCGCGAGA 36689
Qy 1717 uSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnVa 1737
Db 36690 G-----CGACCGCGCGCCAGCC----- 36708
Qy 1737 lProHisLeuProValLeuValProThrProGlyThrProAlaThrAlaMetAspAr 1757
Db 36709 -----GTGGAGTTCTCCGGGTCTCTAGCTCGAAGACGTCTCTCGCGCGCG 36755
Qy 1757 gLeuAlaTyrLeuProThrAlaProGlnPro-----Ph 1768
Db 36756 AGGGCAGTCTTTCGCGCGCGCGCCACCATGTCATGAGGTGATCGAGTCTCCG 36815
Qy 1768 eSerSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysPr 1788
Db 36816 CCGCGTTCGAAGAAGCGTGTCTCCAGCTGACCC---GCTCCAGCGCGACCGCTCCGCG 36872
Qy 1788 oThrThrThrSerSerSerGluArg-----GluArgAspArgAs 1801
Db 36873 AACAGCGCGCACAGCACTGTTTCGCTCGCTGCTCGCGCGCGTCTCGCGCGCGTCCCGCA 36932
Qy 1801 pArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrVa 1821
Db 36933 CGCGATCGGGTCCGGCAGCGCTCGCTCGATCTTCCTCCGTTCCGGGTGGTGGGCGAGC 36992
Qy 1821 lGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerSerGlySerSer 1841
Db 36993 GCTTCCAGCACCA-----CCAGCGCGCGGACCATGTGTCTGGGGCAGCGCGAGCG 37043
Qy 1841 rGlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGln----- 1859
Db 37044 AGGTACTCCGTACCGCGTCCGATCGACCTCCGTTCGGCGACACATGGGCGAGAGCGG 37103
Qy 1860 -----HisSerProIleSerProArgThrGlnAspAlaLe 1871
```

Db	37104	TCGTCGGGAGCGTCAACAGCGCTTGAGCGACCTGCGGGTGGCGGACGACGCGACTCG	37163	Db	38136	ACGAACGCGCTCCGCGCAGCGGACACCCCGACATCCCGACACACACGCGCG	38190
Qy	1871	uGlnGlnArgProSerValLeuHisAlaThrGlyMetLysGlyLeileThrAlaValG1	1891	Qy	2120	uGlnThrAlaPro-----GlyValLysGlyHisGlnArgValThrLeuAlaGlnHi	2138
Db	37164	ATCTCGCGGGCT-----CGATCCGAACCCAGGAGTTCACCTGGTCAATCG	37211	Db	38191	-----ACCGCCCCACCGGAGGCCACCGCGG-----	38217
Qy	1891	uProSerLysProThrValLeuArgSerThrSerThrSerPro-----	1906	Qy	2138	sileSerGluValIleThrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaPr	2158
Db	37212	GCCGCGCCGCGCAACACCACTCACCGCTGCTCGTCAGTGGGCCAGGTCTCTGTCCGG	37271	Db	38218	-----ACTCCGCTCCGACCCCGCGGCGCTCAGCAGC	38249
Qy	1907	-----ValArgProAlaAlaThrPh	1913	Qy	2158	oLeuProLa-----ProLeuTyrSerPheProGlyAlaSerCysProValLeuAsple	2176
Db	37272	TACATCCGCTCACCGGGGACAGGAGTCCGCGACGATCGTTCCCGCTCAGCGCGCA	37331	Db	38250	GCGGCGACCGCAGACGCTGTTCCAGAGCTCCCGAAGTTCAGCT-----CGCG	38300
Qy	1913	eProProAlaThrHisCysProLeu-----	1921	Qy	2176	uArgArgProProSerAspLeuTyrLeuProProAspHisGlyAlaProAlaArgG1	2196
Db	37332	CGGCCAGGTAGCCACCGGCAACCGGATCCCGACACGCTAGAGTTCACCGGTGACCCCG	37391	Db	38301	TCGCGCTCCACCA-----CCGCGCTCTCGTGGCGCGCTCTCTGAAGCACC	38345
Qy	1922	-GlyGlyThrLeuAspGlyValTyrProThrLeuMetGlu-----ProValLeuLeuPr	1939	Qy	2196	ySerProHisSerGluGlyLysArgSerProGluProAsnLysThrSerValLeuG1	2216
Db	37392	GGCGGCAACCGCGGAGGAACGCGTCCAGTACGTGGCAGCGGCTTCCGGATCGGCTCC	37451	Db	38345	-GCGCGCTCACCGGCGCGCGCCCTTCGCGACCGCCACCAACCGGAGGGTCGGCG	38453
Qy	1939	oLysGluAlaProArg-----	1944	Qy	2232	rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh	2252
Db	37452	CCGATCGGGACCACTCCCGCGCGCGCGTCAACGGCATGCTCATCGCGGCACACACCGTG	37511	Db	38394	CGCGCCACGACCGGCGGAGGACTCCCTCGCGCGCGCGCGCGCGCGCG-----	38394
Qy	1945	-----ValAlaArgProGluArgProArgAl	1953	Qy	2216	yGlyGlyGluAspGlyLeuProVal-----SerProPro-----GluGlyMetTh	2232
Db	37512	GCCTCGGTCGCGCGGTAGCGGTTCATCAGCGCTGCTCCCTCGGACCGCGTCCGCGAGG	37571	Db	38395	-GCGCGCTCACCGGCGCGCGCGCCCTTCGCGACCGCCACCAACCGGAGGGTCGGCG	38453
Qy	1953	aAspThrGlyHisAlaPheLeuAlaLysProProAla-----	1966	Qy	2232	rGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnTh	2252
Db	37572	CCCGCGGGGAGGCGCTCCCTCGCCACCAACCGCTCTCCAGGCTCTCCGCGCAGCAGCTCG	37631	Db	38454	GCCACCGCGCGGCGCGCGGCGGACCACTCCAGCGTGCAGTCCG-----CGCGGT	38507
Qy	1966	gSerGlyLeuGluProAlaSerProSerLysGly-----	1981	Qy	2252	rGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePh	2272
Db	37632	GCGTGGCGGACCGACGCGGCGGACCGTCACTGGTGCAGTCCCAACCGCTCCAGTGCC	37691	Db	38508	TCCTCGATCTCGCGTCCGCGTGGTAGCGAGCTGCATCTCAGGTGCTCCCGCGGATGATG	38567
Qy	1981	oArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAs	2001	Qy	2272	eserLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnLysLeuLeuAsnLys	2292
Db	37692	TCGCCACGACACACCGCGCGCGAGTCTCTGCTCGGTG-----	37731	Db	38568	CGACGCTCAGCGGATAGTGGTCCCTCGACCGTCCCGAGATAGTAGTGTGGCGG	38627
Qy	2001	nLeuAlaProHisAlaSerProAsp-----ProAlaProProAlaSerAl	2018	Qy	2292	sLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrG1	2312
Db	37732	-----CCATCACCGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	37784	Db	38628	AGTCCCGCGGTAGTCTTCGAGAGCAGCAGCGGTGCAGACTTCGCGCGCGCGCGCGG	38687
Qy	2018	aSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLe	2038	Qy	2312	uIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaVa	2332
Db	37785	GCGTCGAAGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	37844	Db	38688	ACCTTCGTGATCTCCGCGAGTCCAGGTGCTGATCGGGATCAGGCGCGCGCTCCCGCGG	38747
Qy	2038	uArgSerLeu-----GlyTyrHisGlySerSerTyrSerProGluG1	2052	Qy	2332	lGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetG1	2352
Db	37845	TCGATCTCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	37904	Db	38748	TCGAGACCGGTACAGCATCGGCGA-----CGTCTCGCGCGCGTGCAGCGGACACCGG	38801
Qy	2052	yValGluProValSerProVal-----SerSerProSe	2063	Qy	2352	yLysTyr-----AspGlnTrpGluGluSerProProLeuSerAlaAsnAl	2367
Db	37905	GTTC-----CCGTGACCGCGAGTATAGATACATACGCCAGCTCGTGGCGCCACCGC	37961	Db	38802	ACCGGACCGGTTCATCATGAGCCCGGGCATCGACTCCACCCCGGAGAGTCCCGCGG	38861
Qy	2063	rLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuG1	2083	Qy	2367	aPheAsn-----ProLeuAsnAlaSerAlaSerLeuProAlaAl	2380
Db	37962	GGCAACTCCGCATCCGGGACCGCGGAG-----GATCGACCTCGTGCATCACCGCGAC	38015	Db	38862	CGGACGCGACCGGTGTGGCGAACACCACTCGGTGCGCGCGCGCGCGCGCGCGCG	38921
Qy	2083	uGlyGluLeuArgProLysGlnProGlyProVal-----	2094	Qy	2380	aMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyG1	2400
Db	38016	CGGTCCGCGAATCTCTCGGCGACCGCGCTCCCGTCCGCTCCGCTCCGCTCCGCTCCG	38075	Db	38922	ATCGCCC-----ACGCGCTTCGACCGCGTGTTC	38951
Qy	2095	-----LysLeuGlyGlyGluAla-----	2102	Qy	2400	yGlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaPro--	2419
Db	38076	GGATCCGCGTCCCGGACGAGGCTCAGCGCTCGCGCGGATACGCGGATCCACCGGA	38135	Db	38952	AGCGTCAGCCCGGTTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	39011
Qy	2102	sLeuProHisLeuArgProLeu-----ProGluSerGlnProSerSerProLeuLe	2120	Qy	2420	-----GlyLeuAlaSerGlyArgProProSerValSerValHisSerG1	2436
Db	38076	GGATCCGCGTCCCGGACGAGGCTCAGCGCTCGCGCGGATACGCGGATCCACCGGA	38135	Db	39012	TAGCGATGCGGTACTCGTATCCGCTCCCGCGCGGCTCTTCGCGATCGAGCGGTGCTCG	39071
Qy	2120	sLeuProHisLeuArgProLeu-----ProGluSerGlnProSerSerProLeuLe	2120	Qy	2436	u-----GlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpG1	2451
				Qy	39072	TCGAGCGCGGACAGCTCCGCGTCCGAGGACGCGGCTCGCGGCTCTTGCTCCCGGACCC	39131


```

/notes="Rv0588, (MTCY19H5.34c), len: 295 aa. yrbE2B,
hypothetical unknown integral membrane protein, part of
mce2 operon and member of YrbE family (see citations
below), highly similar to Mycobacterium tuberculosis
proteins O07413|Rv0168|MTCI28.08|yrbE1B (289 aa);
O53966|Rv1965|MTV051.03|yrbE3B (271 aa); etc. Also highly
similar to conserved hypothetical integral membrane
proteins of the yrbE type, e.g. P45392|YRBE_ECOLI
hypothetical 27.9 kDa protein from Escherichia coli (260
aa), FASTA scores: opt: 232, E(): 8.4e-08, (22.1 %
identity in 267 aa overlap); P45030|YRBE_HAEN|H11086
hypothetical protei from Haemophilus influenzae (261 aa),
FASTA scores: opt: 234, E(): 6.3e-08, (24.2 % identity in
215 aa overlap); NP_302655.1|NC_002677 conserved membrane
protein from Mycobacterium leprae (289 aa); etc."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN
YRBE2B"
/protein_id="CAB09950.1"
/db_xref="GI:2222770"
/translacion="MVESSTASAAVLRARYPRTAASLDTRYGGTARRLERTGTFARF
TRISVQIGHALRRYRRETRLVAEIGMCTGMAVVGTVIIGFTVLSGSLIAIQG
FASIGNIGVGAFTGFFAALANTVRAPIVSGVALAATVAGATAQIAGMRISSEIDAL
EVMGKISFLNSTRILGGVLVIMPLVALDINAFITGCGQVTVTFVFGQSGNTYEHYFR
TFLRPEDVGSVVEVETIAVVMITHCYGYTASGPGVGVQAVGRSMRPSLVSVVVV
VLLAELALYGVDPNFLTIV"
2548..3762
/gene="mce2A"
/locus_tag="Rv0589"
2548..3762
/gene="mce2A"
/locus_tag="Rv0589"
/function="UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST
CELL INVASION."
/standard_name="mce2"
/notes="Rv0589, (MTCY19H5.33c), len: 404 aa. mce2A; belongs
to 24-membered Mycobacterium tuberculosis Mce protein
family (see citations below), highly similar to
Mycobacterium tuberculosis proteins
P72013|MCE1|Rv0169|MTCI28.09|mce1A (454 aa);
O53967|MCE3|Rv1966|MTV051.04|mce3A (425 aa); etc. Also
highly similar to others e.g.
AAD52105.1|AF113402.1|AF113402 mycobacterial cell entry
protein from Mycobacterium bovis BCG (454 aa);
NP_302656.1|NC_002677 putative cell invasion protein from
Mycobacterium leprae (441 aa); CAC12798.1|AL445327
Mycobacterium leprae (441 aa); CAC12798.1|AL445327
putative secreted protein from Streptomyces coelicolor
(418 aa); etc. Also highly similar, but longer 21 aa, to
P72013|CAA50257.1|X70901|MTCI28.08 Mcep protein from
Mycobacterium tuberculosis (432 aa), FASTA scores: opt:
1324, E(): 0, (62.6 % identity in 436 aa overlap). Contains
a possible N-terminal signal or anchor sequence. Note that
previously known as mce2."
/codon_start=1
/transl_table=11
/product="MCE-FAMILY PROTEIN MCE2A"
/protein_id="CAE5301.1"
/db_xref="GI:38490208"
/translacion="MPTLVTRKNRRRAWLYVEGVLLLVGALVLVLYKQFGEPTPKT
ELTVAFRAGLVMEAGSKVTYNGVEIGRVGSISEIERDGRPAAKLVLNPNRYLSLIP
VNVVADIEATLFGNKYVALSPKIPQOQRISSHDVIDGVSTVFETNLFETITSIAE
KYVDFELNATLSAVALDGLGKFESEIVNGNQLAOLNPLPOLGVDVRRLLADLGE
VYVDASPDWVSFLQNALITPARTLTSQQQRDLDAALLAATGAGNTGEDVFARGGPVLYA
AADLVPTALLDITYSPLEFCMIRNFHDAKPVADAVGNGVSLAAGTILGAPNPVYV
PNILPRVNAHGGPGRGPCWCQTITRELMPAYLVMDTGLASLAPYNHVELGPMFTEYV
WGRQYGEINTNP"
3759..4586
/gene="mce2B"
/locus_tag="Rv0590"
3759..4586
/gene="mce2B"
/locus_tag="Rv0590"

```

```

Alignment Scores:
Pred. No.: 9.34e-05 Length: 349564
Score: 689.00 Matches: 526
Percent Similarity: 33.44% Conservative: 207
Best Local Similarity: 24.00% Mismatches: 872
Query Match: 5.21% Indels: 599
DB: 1 Gaps: 91

US-09-522-753-5 (1-2517) x BX842574 (1-349564)
QY 546 LeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysGluAlaValAla 565
Db 157203 GTTCGTCGAAGGTCATGGCGGTGCGGTCACGGCTTCGACGAATGCAATGCTCTGCG 157144
QY 566 SerLysGlyArgLysThrAlaAsnSer-----GlnGlyArg--- 578
Db 157143 TCGCGGTGACGAATCTCGAAGATCGCTCGATTGCTGCCAGCGGAGGAACGACGCCCA 157084
QY 579 -----LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSer 592
Db 157083 TACCAGGTTGGCACTGTGAAGGGTCTGACGCCGAGCAGTCGATCGAACCGAGGCGC 157024
QY 593 Glu-----GluAlaIleThrProGlnGln----- 600
Db 157023 GACGGTTCGGTGATTCCGGGTGTCACCGCGGTGTGCAGCCAGCACGCTTGGCGTCCAG 156964
QY 601 -----SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrp 615
Db 156963 AAGGAACATTCGCTTCGCGCTATTGGGGGTACTTCCGGGAGTTTCAGTTCCAGGCC 156904
QY 616 ThrGluGluGluMetGluThrAla----- 623
Db 156903 TCGTCGAGCATTTCGACACACCTCTATTTCGAGGTTCAGTTCGATCGCTCGCGCGGACCG 156844
QY 624 -----LysLysGlyLeuLeuGluHisGlyArg-----AsnTrpSerAla 636
Db 156843 GTGCCCGCGTCGAAACACCGGACCGGTC-----GGCGCGCGCGCGCGCGCGCGCG 156790
QY 637 Ile----- 637
Db 156789 GCGAATCTCCGCGCGAAGGGCGTCTCGATCACAGCGCCCGAGTTCACCGCTCGCGG 156730
QY 638 AlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLys 657
Db 156729 GCCCGCGGTTTGGCGCGCAGCATATTTC----- 156700
QY 658 LysArgGlnAsnLeuAspGluIleLeu-----GlnGlnHis 669
Db 156699 -----TCGAGAGATTGACACCGGTGGTGGCATGATGCACACGTAACGCATCAACAGCAT 156646
QY 670 LysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaProAlaAla 689
Db 156645 CATCTGGCGTC-GAGCGCGCGAAGCGATGCTGTCAACCGCGCGCGCGCGCGCGCGCG 156587
QY 690 AlaSerGluGluAlaAlaPheProValValGluAspGluMetGluAlaSerGly 709
Db 156586 CGGTACATCAAGCGGTTTCAGCCCGCTTCACCGCAGCAGCGCGCGCGCGCGCGCGCG 156527
QY 710 ValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsn 729
Db 156526 GCCCGCTGG----- 156518
QY 730 GluValProArgGlyGluCysSer-----GlyProAlaThrValAsnAsnSer 745
Db 156517 -----GCCCGGTGCACCCCGGATTTTCCCGCGGTATACCGCTGTATTACCGCGTT 156470
QY 746 SerAspThrGluSerIleProSerProHisThrGluAlaAla----- 759
Db 156469 GCCCGCGGTTCGCCCGCTTTCGCGATCAGCAGCGCGCTGCCCGCGCGCGCGCTTCCGCGC 156410
QY 760 -----LysAspThrGlyGlnAsnGly-----ProLysProProAlaThrLeuGlyAl 775
Db 760 -----LysAspThrGlyGlnAsnGly-----ProLysProProAlaThrLeuGlyAl 775

```



```
Db 156409 GATGTGAAACACCGGACAAACCGGTGGCGCCCGCGCGCGTCACTCCGAGACC 156350
Qy 775 aAspGlyPro-----ProProGlyProProThrProPro-----ArgArgThrSerAr 791
Db 156349 GAACGCTCCGACACTTGGCGCGCGCGCGCGTCCCGCGGTGGCGCGCGTGGCGGA 156290
Qy 791 gAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProAl 811
Db 156289 GCGCGCGCGCGCGCGCGCGCG-----CGGAGCGCGACGAAACCGCGTCCGCGCGCG 156233
Qy 811 aProProSerProSerAlaPro-----ProProValProLysGluGluLy 827
Db 156232 CCCCGCGCGCGCGCGCGTGTGCTGAACCCCGCGTCCCGCGCGCGCGCGCG 156173
Qy 827 sGluGluGluThrAlaAlaProProValGluGluGluGluGluGluGluGluGlu 847
Db 156172 GCCAAGAGCAGCGCTCCGCGCGCGCGCGATG-----CGCC 156137
Qy 847 aAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysTh 867
Db 156136 GACCACCGCGCGCGCGTCAAG-GCTGGAACCGCG----- 156104
Qy 867 rGluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 887
Db 156103 -----GCTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 156051
Qy 887 aGluGluAlaLeuLysAlaGluLysLysGluGlySerGlyArgAlaThrAlaLy 907
Db 156050 CGGCCCGCGCGCGCGCGTCCGGAACTCCGCGCGCGGTGTATGGCTCCGCGCGCGCGG 155991
Qy 907 sSerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAlaAspGluValAs 927
Db 155990 TGCGCGCGCGCGCGCGCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGG 155931
Qy 927 pGluAlaGluGlyClyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThPr 947
Db 155930 TGCTTCCCGCGCGCGCGCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGTTCACCTG 155871
Qy 947 oThrGlyAspProArgAlaAlaAlaSerProGlnLysProLeuAspLeuLysGlu 967
Db 155870 CCCCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCC-- 155814
Qy 967 sGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArg 987
Db 155813 -CCCGGTG-CGGAACAGCGCGCGTGA-----CGCGCGCGCGCG 155777
Qy 987 uAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsnLeuGlnPr 1007
Db 155776 GCCATTGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155737
Qy 1007 oGluSerAspAlaProGlnProGlnProGlySerSerProArgGlyLysSerArgSerProAl 1027
Db 155736 -----GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155696
Qy 1027 aProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProPr 1047
Db 155695 CCGCCATTGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAA 155636
Qy 1047 oCysTrpThrSerGlyLeuProPheProValProProArgGluValLleLysAlaSerPr 1067
Db 155635 GAGCAAGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155576
Qy 1067 oHisAlaProAspProSerAlaPheSerTrpAlaProProGlyHisProLeuProLeuG 1087
Db 155575 GCTGCCCGCGGTGGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155516
Qy 1087 y-----LeuHisAspThrAlaArgProValLeuLeuProArgProThrIleSerAsnPr 1105
Db 155515 CCGCGCGTTCGACCGCGCGTTCCTTCCTCCCGCGCGTTCCTCCCGTTCGCGAACAACCA 155456
Qy 1105 oProProLeuLeuSerSerAlaLysHisProSerValLeuGluArgGlnLleGlyAlaI 1125
Db 155455 GCGCGCGCGTCCGCGCGTCCGCGCGCGTCCGCGCGTCCGCGCGTCCGCGCGT 155406
```

```
Qy 1125 eSerGln-----GlyMetSerValGlnLeuHisValProTyrSerGluHisAl 1141
Db 155405 TGCGGATCAGCGGACGCCCGGTGAGCGCTCTGGAGAGGGTCTTCCACCATATTGATCAGT 155346
Qy 1141 a----- 1141
Db 155345 CCTGCTCAGGGTGTTCAGGGGCGAGGTGTGGCGGGGCGGTAAAGCCGCTCCAGGCGCA 155286
Qy 1142 -LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLe 1161
Db 155285 ACAGACCCCGCGCGGTACCGCGCATGCGCGCTTGCACAGGTGCGCGCGTCCCGCGC 155226
Qy 1161 uAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProPr 1181
Db 155225 TGCGCGGTTCGCGCGTTCGCGATCAACACCGCGGTGCGCGCGCGCGCGCGCGCGCAC 155166
Qy 1181 oGluSerLeuGlyValProThrAla-----GlnGluAlaSerValLeuArgGlyThrAlaLe 1200
Db 155165 C-----GTCCCGCGCGGTGCTGCGCGCGCGGTTCGCGCGCGGTTCGCGCGGT 155118
Qy 1200 uGlySerValProGlyGlySerIleThrLys-----GlyIleProSerThrArgVa 1217
Db 155117 TGCGGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTCCTCCGAGCGTTC 155058
Qy 1217 lProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspVa 1237
Db 155057 TAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155003
Qy 1237 lLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspAr 1257
Db 155002 -----GCCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154971
Qy 1257 gGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLys----- 1273
Db 154970 TGCGACTAGCGCGCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 154911
Qy 1274 ---LysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysG 1292
Db 154910 TGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 154852
Qy 1292 uAspGlyArgSerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAs 1312
Db 154851 CTTCCGCGCGCAAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC-- 154794
Qy 1312 pMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetG 1332
Db 154793 ---CGCGGTTCGAGAATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154738
Qy 1332 yArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisIleAr 1352
Db 154737 GCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 154684
Qy 1352 gGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuAr 1372
Db 154683 G-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154645
Qy 1372 gArgGluAlaLysLeuLysArgGluGlyThrProProProProProProProProPro 1392
Db 154644 CCGCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154594
Qy 1392 pLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisG 1412
Db 154593 -----CGGTG----- 154588
Qy 1412 uGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluG 1432
Db 154587 -----CCGATCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154562
Qy 1432 uLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlySerIleTh 1452
Db 154561 GTTGGCTCCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154519
```

```
QY 1452 rGlnGlyThrProLeu-----LysTyrAspTh 1461
Db 154518 -ACGGGCGCCGCGTGGCCGACGAATTCGGCGTTGATCGGGCCAGCGCGGACAC 154460
QY 1461 rGlyAlaSerThrThrGlySer-LysLysHisAspValArgSerLeuIleGlySerProG 1481
Db 154459 GCGCGCGCGCTCGCGCGCGCATACGCGCTCCGCGCGGCTCAGGGCTGACAGAACTG 154400
QY 1481 lYArgThrPheProProValHisProLeuAspValMetalAspAlaArgAlaLeuGluA 1501
Db 154399 GCGGTAAACGCGCGCG-----CCTGTG----- 154377
QY 1501 rGalaCysTyrGluGluSerLeuLysSerArgProGly-----ThrAlaSerS 1517
Db 154376 -----CGTCCGCGCGCTGGTAGCGCTGCGCTTGGGACCGACCAACCGCGC 154331
QY 1517 erSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProA 1537
Db 154330 AATCGCTGTGACACCTCATCGGACCGCGCGGTCAACAATCCGTCGTGTTGACGCGCG- 154272
QY 1537 rglnSerProLeuThrTyrGluAspHisGlyAlaPropheAlaGlyHisLeuPro----- 1555
Db 154271 -----CGCGCGCATATAGCGAGCT 154253
QY 1556 -----ArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluG 1572
Db 154252 AATCGCGGACCGGATCCCGCGCAAAATCCGCTGCGCGCGCCACACGACTCTCCGCGGAAC 154193
QY 1572 lySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSer----- 1587
Db 154192 CATCACCATGACATCGATCGCTCCCAA-----CCCTCATCGCGCTGTGTCGCGACTA 154139
QY 1588 -----ThrProArgGluIleAla-----LysS 1595
Db 154138 CCAGCGCACACCGGTACGGGTAGGAATTAATTCAGAAATCAACGATATCGCGATCCGG 154079
QY 1595 erProHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuL 1615
Db 154078 CACCGCATGACGGCTTCCACACGTTTCGCAG-----GGCCAGCTGTGGACGGC 154028
QY 1615 euArgGlyValSerGlyValAspLeuTyrArgSerHisIlePro-----LeuAlaP 1632
Db 154027 TCAGGGTCGCGCGCGGCTTCCGACAAACGGCGCGGTGTCGCGGACAGGGCGCGCG 153968
QY 1632 heAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyrTyrLeuP 1652
Db 153967 CGAAACCCGCA-----CCGCGCGGATAGCGCTGTCACGACGACCGCGC----- 153924
QY 1652 roArgHisLeuAlaProAsnProThrTyrProHisLeuTyr-ProProTyrLeuIleArg 1671
Db 153923 --ACGGGCTCTGACAAAGCGCTCAGCGCGCGGGATCTTCGCGAGATACCTGGCGCG 153866
QY 1672 GlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThr 1691
Db 153865 CG-GCCGCCATTTTGGCGGCGCG-----TCC 153840
QY 1692 SerGlnGlnMetHisHisAsnThrAlaThrAlaMetalaglnArgAlaAspMetLeuArg 1711
Db 153839 GCAGCGCATCTGCATCACCGCGCGAGTGGTGACGCCCGGAGCCATCCGCTGCGCGCGCG 153780
QY 1712 GlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly-Pro-----Ar 1729
Db 153779 CTACGGCAACCCGTTGAGCCGCTTGGCCCGCAGCAGCAACCGCGCGCGCGCGGTTC 153720
QY 1729 gGlyIleLeAspLeuSerGlnValProHisLeuProValLeuValProProThrProGl 1749
Db 153719 GGGGCTCGCGCGTGGCGCGGTGCGCGCGTTCGCGCGCTG-----CCGCGGTGCGCAT 153663
QY 1749 yThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSe 1769
Db 153662 CAGCCCGCGCGTGGCG-----CCGAGCCACCGCGCGCG----- 153629
QY 1769 rSerArgHisSerSerProLeuSerProGlyGlyProThr-His-----LeuT 1786
```

```
153628 -CCGAAGCCACGCGCGCATTCGCGCGGCCCCACGACCCACCGCGCGGTGCGGAA 153570
1786 hrLysProThr-----ThrThrSerSerGluArgGluArgAspA 1800
153569 CCGCGCGAACCGCGCGCGGCCCGCGCGTGGACAAACAGCAGCCGCGGTTCGCCACCGCGCC 153510
1800 rGAspArgGluArgAsp-----ArgAspArgGluArgGluLysSerIleLeuThrS 1817
153509 GCGTCCCCCAGGAGAACCAATCCGCGCGCGCGCGCGCGCGCGCGCTGCCGACAG 153450
1817 erThrThrThrValGlu-----HisAlaProIleTyrArgProGlyThrGluGln-Ser 1834
153449 CAGCCCTCGCTACCGCGCGCGGCCCGCGTCCACCGGTGTCCAGGTTTCGCGGTCCAC 153390
1835 SerGlySerSerGlySerSer-----GlyGlyGlyGlyGlySerSerSerArgProAla 1852
153389 GCGGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 153330
1853 SerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGln 1872
153329 AACCCTCGCTGCGCATAGTCTGAGCCCGCGCGTCCCGCGTCCCGTCCCGCGCGCAC 153270
1873 GlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPro 1892
153269 GATCAGCCCGCG-----CAGCAGTCCGCGCGTCCCGCGCGC-----CCG 153231
1893 SerLysProThrValLeuArgSerThrSerSerSer-ProValArgProAlaAlaTh 1912
153230 ACCGCGCGCGCGCTC-----GCCACCAAGCGCGGTTCACCGCGCGCGCGCG 153183
1912 rPheProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLe 1932
153182 GCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 153150
1932 uMetGluProValLeuProLysGluAlaProArgValAlaArgProGluArgProAr 1952
153149 GCCTCCCGCGCGTCCGCGCGAGT-----ACGGTGTGCTGCGTCCCGCATCCCGC 153099
1952 gAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAl 1972
153098 GCGCGCGCGTCCCGCGAGC-----AGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 153048
1972 aserSerProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAl 1992
153047 CCGCGCGCACCGCGCGAGTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGTACCC---CACAG 152991
1992 aThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProPr 2012
152990 CCGG---GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152934
2012 oAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSe 2032
152933 GCGCGCGCGTCCCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152906
2032 rIleGlnGluLeuLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGl 2052
152906 ----- 152906
2052 yValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLy 2072
152905 -----CCGGGCGCGCGGTGCG-----CCGCTGCCGACCGCGCGCGCGGTGCGG-- 152861
2072 sHisLeu-GluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProG 2092
152860 -----TTGCCGATCAGCGTCCCGCGTTCAGTGTCTGGTGGCGCGCTGTATCGCGCGCAG 152805
2092 lyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluS 2112
152804 CGCTGT---GTTTCAGGGTGTGACGCGAGAG-GCGCTGGCGCGGGGTGTTGAAGCGGT 152749
2112 erGlnProSerSerProLeuGlnThrAlaProGlyValLysGlyHisGlnArgV 2132
```

```
Db 152748 CGCGCCAGCAGAGCCACTGATCCACCCGCGCGC----- 152711
Qy 2132 alValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHisP 2152
Db 152710 -----GTATCACCGCGGTGCGTTCGGTT-----CCGC 152683
Qy 2152 roGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysP 2172
Db 152682 CGATCGCGCGGTTCGCGGTTCGCGATCAGCACCCCGTTCGCGCGCGCGCGGTAC 152623
Qy 2172 roValLeuAspLeuArgProPro-----SerAspL 2183
Db 152622 CGCGGTCACCGCGCCCTTGCCACCGCGCGCGCGCGCGCGCGCTTACCGATCACCAACC 152563
Qy 2183 euTyrLeuProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyG 2203
Db 152562 CGGCTTGGCGCGCGCGCGCGACACCGCGTTCACCGGTTTCGCGCAACCGCGCGCGCC 152503
Qy 2203 lyLysArgSerProGluProAsnLysThrSerVal-----LeuGlyGlyGlyGluA 2220
Db 152502 CGCGGACCCACCGGAGCCCAACACAGCC-GGCGTTCACCGCGCGCGCGCGCTCCCGC 152444
Qy 2220 spGlyIle---GluProValSerProProGluGlyMetThrGluProGlyHisSerArgS 2239
Db 152443 TCGGTATCGCGGAACCCCGCGTCCCGCGGATCCGCGGTTCGCGAAGACACACCGGTG 152384
Qy 2239 erAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSer----- 2255
Db 152383 TTGCGCGCGCGCCC-----ACGCGCCACCGTCCACCACT 152348
Qy 2256 -----ArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheS 2273
Db 152347 AGGCGGCTCGCGCGATCCACCGGTTCGCGCGCGCACCGAAGACAGCCCGCGGTTCGC 152288
Qy 2273 erLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLysL 2293
Db 152287 CCAGCCCGACCGACCCCGC-----GTTGTTGAGA 152258
Qy 2293 ysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly----- 2310
Db 152257 AAGCCAGTTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 152198
Qy 2311 --ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerG 2330
Db 152197 CCACCGGCTCGCGAGTTCGCCCGACCCACCGGTGACGGCGC-CTAGCCCGCGCGCGCC 152139
Qy 2330 lnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaL 2350
Db 152138 GCGGTGCCCGCGCAC----- 152122
Qy 2350 euMetGlyLysTyrAspGlnTyrGluGluSerProProLeuSer-AlaAsnAlaPheAsn 2369
Db 152121 -----CCACMGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 152097
Qy 2370 ProLeuAsnAlaSerAlaSerLeuPro-----Ala 2379
Db 152096 CCGCGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 152037
Qy 2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu----- 2394
Db 152036 CAACAACCGACCACTGATCCACCGCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGC 151977
Qy 2395 -----ThrSerProGlyGly-----GlyGlyLysAlaLysValSer 2406
Db 151976 CCGCGCGCGCGCACCGCGCGCGGTTCGCGAACAACCGCGCGCGCGCGCGCGCGCGCGCGC 151917
Qy 2407 -GlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspAr 2426
Db 151916 GGGCAGCGCGCGAGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 151860
Qy 2426 gProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuTh 2446
Db 151859 CCGCGCGCGCGCACCATTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 151815
```

```
Qy 2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhe---ProTyrAs 2465
Db 151814 GCGTTCGCGATCAACGGCGCGCGCGTGGCGGACACACTTTTGGCATTTGACGGCGGCCAG 151755
Qy 2465 nProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProGlyLeuPr 2485
Db 151754 CAGTCTCTGTCACGCTCTCGCGCGGTTCGCGGCGGTTGGCCAGCACCGCGCGCGGC 151696
Qy 2485 oAlaGlySerGlyPro 2490
Db 151695 CGTGGCGTAGGCCCC 151680

RESULT 76
E38020
LOCUS E38020 Avermectin aglycon synthase gene. linear PAT 31-JAN-2002
DEFINITION E38020
ACCESSION E38020
VERSION E38020.1 GI:18626909
KEYWORDS JP 200245457-A/1.
SOURCE Streptomyces avermitilis
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 30690)
AUTHORS Omura, S. and Ikeda, H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 200245457-A 1 12-SEP-2000;
THE KITASATO INSTITUTE
COMMENT OS Streptomyces avermitilis
PN JP 200245457-A/1
PD 12-SEP-2000
PR 24-FEB-1999 JP 1999046961
PI SATOSHI OMURA, HARUO IKEDA
PC C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
C12Q1/68//C07D493/22,
PC (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
FH Key Location/Qualifiers
FT CDS (1)-(11916)
FT CDS (11971)-(30687).
FEATURES
source
1..30690
/organism="Streptomyces avermitilis"
/mol_type="genomic DNA"
/db_xref="taxon:33903"

ORIGIN

Alignment Scores:
Pred. No.: 1.59e-05 Length: 30690
Score: 685.00 Matches: 704
Percent Similarity: 30.18% Conservative: 300
Best Local Similarity: 21.16% Mismatches: 1193
Query Match: 5.18% Indels: 1144
DB: 6 Gaps: 148

US-09-522-753-5 (1-2517) x E38020 (1-30690)

Qy 44 HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArg 63
Db 11193 CATCATCGTTCGGTCTGCATCCATCCATGACCCCGGA-----CGC 11234
Qy 64 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHis 83
Db 11235 ACTGGCCCTTTCGATCGCGC-----CCTGGCTCTGGACCGCGCGTCTGTGTGCC 11285
Qy 84 LeuArgProGluSerHisSerTyr-LeuProGluLeuGlyLysSerGluMetGluPheIl 103
Db 11286 CGCGGACCTCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 11336
Qy 103 eGluSerLysArgProArgLeuGluLeuLeu-Leu-ProAspProLeuLeuArg---ProSerP 122
Db 11337 GCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 11396
```


Db 13363 CCGCCACACATCCCGTTCAGACACACCCCGCGACGATGTCGCGGAGATACGCGCC 13422
Qy 685 ys-AlaProAlaAlaSerGluGluAlaAlaA- - - - -PheProPro 698
Db 13423 GACGAGATCCGGTAGTGGGATGAGCTCTCCCGGAGTCCAGGGGTGTGCGGTGG 13482
Qy 699 ValValGluAspGluGluMetGluAlaSerGlyValSerGlyAenGluGluMetVal 718
Db 13483 CTGGTGTGGCCAAAGTCGACCGGCC- - - - -CTGCGC 13515
Qy 719 GluGluAlaGluAlaLeuHisAla-SerGlyAenGluValPro- - - - -ArgGlyGluCy 736
Db 13516 GCCGAGGCCAGGCCCTGACGACCCATCCACCGACCCCGGCTCGAGCTCGCGGAC 13575
Qy 736 sSerGly- - - - -ProAlaThrValAsnAsnSerSerAspThrGluSerI1 751
Db 13576 GTCCGGTACACCTCGCCCGCCGCGCGCTGTTCGACACCCCGCCCGCCCTCATGCC 13635
Qy 751 ePro- - - - -SerProHisThrGluAlaAlaLysAspThrGlyGlnAenGlyProLysPr 769
Db 13636 GCCGACCGGACACCTTCCTCAAGCACTCCAGGCACTCGCGCGAGGACCCACCCC 13695
Qy 769 oProAlaThrLeuGlyAlaAep- - - - -GlyProProProGlyProPro- - - - - 783
Db 13696 GCCGTATCTCCAGCAGCGCCCGCGGCGGACCGGCGACCGGGGAGCGCGAGAAAGACC 13755
Qy 784 - - - - -ThrProProArgThrSerArgAlaProIleGluProThr- - - - - 797
Db 13756 GCATTATCTCTCCGAGACGGGACCCCAACGCCCGCGGATGGCCACCGGCTCTACAC 13815
Qy 798 - - - - -ProAlaSerGluAlaThrGlyAlaPro- - - - - 806
Db 13816 ACCACCCCGTCTCGCGCGCGCATCAACGACATCTGCACCCACCTCGACCCCACTC 13875
Qy 807 -ThrProProAlaProProSerProSerAlaProProValProValProLysGluI 826
Db 13876 GACCACCCCTCTCTCCCGCTCTCACCAGGACCCCAACACCCAGGACACCCACCCCTC 13935
Qy 826 uLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProPr 846
Db 13936 GAAGAGCGCGCGCTCTCCAGCAGACCCCGCTACGCGCGCGCGCTCTTCGCGCTTC 13995
Qy 846 oAlaAlaGluGluLeuAlaValAsp- - - - -ThrGlyLysAlaGluGluProValLysSerGl 865
Db 13996 CAGTTCGCCCTCCACCGCTCTCACCAGCGCTACCAATCACCACCCCGCTACTACGCC 14055
Qy 865 uCyThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaGluAl 885
Db 14056 GGACACTCTCTCGCGGAAATCACCAGCGGCCACCTCGCGGATCTCTACCTCACCAG 14115
Qy 885 aThrAlaGluGlyAlaLeuLysAlaGluLysLys- - - - - 896
Db 14116 GCCACACCTCTACCCAAACCGCGCCACCTCATGAAACATGCCCGCGGACCATG 14175
Qy 897 -GluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
Db 14176 ACCACCTCCACACCCCGCCCAACATCACCACCTCACCAGCGCCAGGAAACGAC 14235
Qy 916 pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnAr 936
Db 14236 CTCGCCATCGCGGCATCAACACCCCGCCACCTCTCTCTCATCGCGCA- - - - - 14284
Qy 936 gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAlaAlaSe 956
Db 14285 - - - - -CCCCCAACCGTCCCAACATCACCA- - - - -CCCTCTGCAACAA 14328
Qy 956 rProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaAlaPro- - - - -Pr 975
Db 14329 GGCATCAAAACCAAAACCTCCCAACCAACACCGCTTCCACTCCCGCCACACCAACCCC 14388
Qy 975 oIleGlnValThrLysValHisGluPro- - - - -ProArgGluAspAlaAlaProThrLysPr 994
Db 14389 ATCCTCAACCACTCCACACGACACCCAAACCTCTACCTACCAACCCACCCACCCACCCC 14448
Qy 994 oAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGl 1014
Db 14449 CTCATCACCGCAACACCCCGCAACTCTCTACCCCGCTACTGACCAACAA 14508
Qy 1014 nProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAl 1034
Db 14509 GCCGCAACA- - - - -CGTCTGATCTACGCAACCAACCCCAACCCCTCCACCA 14556
Qy 1034 aPheAlaAlaGluAlaGlnLys- - - - -LeuProGlyAspPro- - - - -ProCy 1048
Db 14557 CACGGCTCTCCACCTATCTGAACTCGAATCGGACCCGCAACACCTCCACCCCTCACCCAC 14616
Qy 1048 sTyrThrSerGlyLeuProPheProValPro- - - - -ProArgGluValI1 1063
Db 14617 GACACCTCTCCCAACACCCCGCAACCTCTACCTCACCACCCCGCCACCCACCCACCCC 14676
Qy 1063 eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyrAlaProProGly- - - - -H 1082
Db 14677 CAAACCCACCTCTCTCAACACCTCGCAAAACACCAACCAACCACTGCGACCCCGCCACCTAC 14736
Qy 1082 isProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrI 1102
Db 14737 ACCACCAACA- - - - -CAACCAACCCCAACCAACCAACCTCGACCTCCCGCAC 14787
Qy 1102 leSerAsnPro- - - - -ProProLeuLysSer- - - - -SerAlaLysHisProSerV 1117
Db 14788 TACCCCTTCAACACCAACCACTCTGCTCTCAACCAACCCCGCAACCGGACCGACCC- - - - -G 14844
Qy 1117 alLeuGluArgGln- - - - - 1122
Db 14845 TCACCGAGCGAGCGCGTAGAAGCCAGCAGCCCAACCCCGCTGGGTGATGTCCTC 14904
Qy 1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGlu-HisAla 1141
Db 14905 GTGGGCAAGTCTCGCAGGAGCGAGACAGAGAGCTGTTCGCGCTGGTGCGACCATGCG 14964
Qy 1142 LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu 1161
Db 14965 GCCGTGTGTGGCCATGCCATCCCGAAGTGTTCGGAACAGAGCCCTTCAAGAG 15024
Qy 1162 AlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro- - - - - 1180
Db 15025 CTGGGTTTGTATCTCTCGCGCAATTCAGCTTCGTAATCGACTGCTGTGACGTTGAC 15084
Qy 1181 - - - - -ProGluSerLeu- - - - -GlyValProThrAlaGlnGlu- - - - -AlaSerVal 1194
Db 15085 CTGCGCTTCCGGCCACCGCTGATCTTCGATTACCCCACTCCGATGGCGCTTTCAGGTT 15144
Qy 1195 LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSer 1214
Db 15145 CTCGGCGCGGATCTGCGGAGCGGACAGAGC- - - - -AGGACACTCTGCTCTCCGCTA 15198
Qy 1215 ThrArgValProSerAspSerAlaIleThr- - - - -TyrArgGly 1227
Db 15199 ACTCGGTTCCTCGCGAGCGGATGCCATCGTCGCGATGGCTCTGTCGTACCCCGT 15258
Qy 1228 SerIleThr- - - - - 1230
Db 15259 GATGTACGAGCGGTGATGATCTCTGCGAGGTGCTGAGTGTGGCCATGACGCGATCGGC 15318
Qy 1230 - - - - - 1230
Db 15319 GGATTCGAGAACCGTGGTGGGACCTCGACACGCTGTACACCCCGGACCCCGGACAC 15378
Qy 1231 HisGlyThr- - - - - 1233
Db 15379 CACGGAACAGCTACACCCCGGCGGGATTCCTTTACGACGCGAGCAATTTTCGATCCC 15438
Qy 1234 - - - - -ProAlaAspValLeu- - - - - 1238
Db 15439 GACTTCTTCGTATCAGTCCGCGTGAGGCACTGCGGATGGACCGGACGCGCTGCTG 15498

QY	1239	-----	-----TyrLysGly	1241	QY	1436	Thr	-----ProGluLeuProLeu	-----	1441
Db	15499	CTGGAACACGGTGGGAGATCGAACACGCGCTGCATCAACCCGACAGCCTCCGTGGC	-----	15558	Db	16500	ACCCGCCACAACTCCCGTCAGACACACCGCCGAGAGACGCGCG	-----	16559	
QY	1242	ThrIleThrArgIleIleGly	-----GluAspSerProSerArg	1254	QY	1442	-----	-----AlaProArgProLeuLysGluGlySerIleThr	1452	
Db	15559	ACACCAACCGCGCTCTTCCGCGGCTGACCTACACGACTACCGCGCGCTTCCCA	-----	15618	Db	16560	CGACGATGTCGGGGGAAGCGCGCGACGACGCGGTACCGCGGGGAAGC	-----	16613	
QY	1255	-----	-----LeuAspArgGlyArgGluAspSerLeuProLysGly	1266	QY	1453	GlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAsp	1472		
Db	15619	GCTCCGCGACGGTTCGAGGGGTATCTCGGCACGGAAGCCGACGATATCGCTCGGT	-----	15678	Db	16614	GACTGGTCTGCTCGCGCAGTCCAGGGGTGTGGCTGGTGTGGCTGGCCAAATCGCA	16673		
QY	1267	HisValIleTyrGlu	-----GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer	1285	QY	1473	ValArgSerLeuIleGlySerProGlyArgThrPheProPro	-----	1486	
Db	15679	CGTGTGCGCTACGCTCTCGGCTCGGAGGTTCGGCCCTCACAGTCGAC	-----	15726	Db	16674	GCGGCGCTGCGGCGCCAGGCGCCCTGCAGCGCCACCTCACCGACCCCGGCT	16733		
QY	1286	ValThrGlnCysSerLysGluAspGlyArgSerSerGlyProHisGluThrAla	1305	QY	1487	-----	-----ValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu	1500		
Db	15727	---ACTGCTGCTCTCGTCCCT--GGTCGCTCTGCA--	-----CCTGCGCTGTGAGCGCT	15776	Db	16734	CGACCTGCGCGACGTGCGGTACACCT	-----CGCCACGCGCG	16772	
QY	1306	AlaProLysArgThrTyr	-----AspMetMetGluGly	1316	QY	1501	ArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSer	-----	1518	
Db	15777	GCGTCCGGGAGGTTCATGCGCTCGGCGGTGGCGTACGCTGATGTCAACCCCGC	-----	15836	Db	16773	CGCG--TGTTGACACCGCGGCCACCTCATCGCGCGCAGCCGACACCTTCTTGCAG	16831		
QY	1317	ArgValGlyArgAlaIleSerSerAla	-----SerIleGluGlyLeu	1330	QY	1519	-----	-----GlyGlySerIleAlaArgGlyAlaProVal	1528	
Db	15837	CGGGTTCGTGAGTTCGCGGACGCGGCTGGCGGTGGACGCGGTGCAAGCGGT	-----	15896	Db	16832	CATCCAGCAGCTCGCGCGAGAACCCACCCCGCGCTCATCCACGACGCGCCCGAC	16891		
QY	1331	MetGly	-----	1332	QY	1529	IleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla	1548		
Db	15897	CTCGGACGCGCTGACGCGACCGGCTGGGTGAGGTGTGGAATGCTGCTGGTGAGCG	-----	15956	Db	16892	GCGGACCGCGGACCGGGAGCGCGAGAAAGACCGCATTCATCTGCTCGGACAGGGCA	16951		
QY	1332	-----	-----	1332	QY	1549	ProPheAlaGlyHisLeuProArg	-----	1556	
Db	15957	GCTGTGCGACGCGCGGCTCGGTACCGAATCTCGCGTGTGTGGCAGTGGCGT	-----	16016	Db	16952	CCACGCGCGCGCATGGCGCCACGCGCTTACCACACCCCGCTTTCGCGCGCGAC	17011		
QY	1333	-----	-----ArgAlaIlePro	1336	QY	1557	-----	-----GlySerProValThr	1561	
Db	16017	CAATCAGGACGCTGCGAGCAACGGGCTGACGCGCCCAACGGGCGCTCCAGGAGGTGT	-----	16076	Db	17012	TCAACGACATCTGACCCACCTCGACCCCGCTCGACCCCGCTCTCTCCCTCTCTCA	17071		
QY	1337	-----	-----ProGluArgHis	1341	QY	1562	MetArgGluProThrProArgLeu	-----	1569	
Db	16077	CATCGCTGCGCTGGCCACACGGGCTGACCCCGCGACGCTGATGCGGTGGAGGC	-----	16136	Db	17072	CCAGGACCCCAACACCCAGGACACCCACCGTCCGACCCCGCTTCCCTCTCTCTCA	17131		
QY	1342	ProHisHisLeuLysGluGlnHisIle	-----ArgGlySerIleThrGlnGlyIlePro	1360	QY	1570	-----	-----GlnGluGlySerLeuSerSerSerLysLysAlaSerGln	1581	
Db	16137	CCACGGCAC	-----CGGCACCATTTTGGGCGACCCGATCGAGGCCCGCCCTCT	16187	Db	17132	AGACCGGTACGCCACCGCGCTCTTCGCTTCCAGGTGCGCTCCACCGCTCTCTCA	17191		
QY	1361	ArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLysArg	1380	QY	1582	AspArgLysLeuThrSerThrProArgGluIle	-----	-----AlaLysSerPro	1596	
Db	16188	CGCCACCTA	-----	16199	Db	17192	CCGACGGCTACCATATCACCCCGCTACTACTCGCGGACACTCTCTCGCGGAATCACC	17251		
QY	1381	GluGlyThrProProProProSerArgAspLeuThrGluAlaTyrLysThrGln	1400	QY	1597	-----	-----HisSerThrValPro--GluHisHisProHisProHisSerPr	1610		
Db	16200	ACAGACCGCCCGCAACGACCGCTGTGGCTCGATGAAGTCGAACATCGGCCA	16259	Db	17252	CGCCCGACCTCGCGGATCTCTACCTTACCGACCGCACCCCTCATCACCACCGCG	17311			
QY	1401	AlaLeuGlyProLeuLysLeuLysProAlaHisGlu	-----	1412	QY	1610	oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLe	1630		
Db	16260	CGCGAGGCTGCGGAGGTGTGGCGGGTTCATCAAGATGTGTGCGCTCGGAAATGG	16319	Db	17312	CCACCTCATGCAACCATGCCCGCGCACCATGACCACTCCACACCCCGCCACC	17371			
QY	1413	-----	-----GlyLeuVal	1415	QY	1630	uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaAlaTyr--	1649		
Db	16320	TCTGTGCGCGGACGTTCATGTGTGATGACCGCTGCGCGCATGTGGACTGTTCGGGG	16379	Db	17372	ACATCACCCACCTCACCGCCCGAAGAACGACCTCGCATCGCGGCATCAACACC	17431			
QY	1416	-----	-----AlaThrValLysGluAla	1421	QY	1650	-TyrLeuProArgHisLeuAla	-----	1656	
Db	16380	GCGGTGACGTGCTACGAGACGCTGCTGCGCGCGGAGGGCGGCTGCGGCG	16439	QY	17432	CGACCTCTCTGTCATCAGCGGCGCACCCCGCACCGTCCAAACATCACCACCTCTGCC	17491			
QY	1422	GlyArgSerIle	-----HisGluIleProArgGluLeuArgHis	1435	QY	1656	-----	-----	1656	
Db	16440	GGCAGAGTGTATCTGCTGCGGTCAGCGGCGCACCAACGCCACGCTCATCTCTGAAAGC	16499	Db	17492	AACAACAAGGATCAAAACCAAAACCTCCCGCACCAAAACGGCTTCCACTCCCCCACA	17551			
QY	1422	GlyArgSerIle	-----HisGluIleProArgGluLeuArgHis	1435	QY	1657	-----	-----TyrProHisLeu-TyrProPro	1668	

[illegible]

Qy	2247	gAspGlyGluInThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerG1	2267	ORGANISM	Streptomyces avermitilis
Db	19406	CTGGGGACCTGGCAAGG---AAACGGAGTCGGGACCTGGTCAAGTCAGCGACACTCTCC	19462	REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Qy	2267	nProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysL	2287	AUTHORS	Streptomycineae; Streptomycetaceae; Streptomyces.
Db	19463	GCCTCCCGCGGA-TGTTCGCCATGCGCGCGAGTTGGGTACACACTGTGACGGCGG	19521	TITLE	1 (bases 1 to 30690)
Qy	2287	sGlnGluLeuAsnLysLeuAsnThrHis-----AsnArgAsnGluProG1	2303	JOURNAL	Endo,H., Yamaguchi,H., Kanda,Y., Hashimoto,S., Omura,S. and
Db	19522	ATCGGAGCGCGCGCGAGTCCTCGTCGCGATATGACGTGGAAGAAATTGGGACCG	19581	COMMENT	Ikeda,H. A method for procuding avermectin derivatives Patent: WO 0162939-A 1 30-AUG-2001; KYOWA HAKKO KOGYO CO LTD, THE KITASATO INSTITUTE, HIROFUMI ENDO, HIROYUKI YAMAGUCHI, YUTAKA KANDA, SHINICHI HASHIMOTO, SATOSHI OMURA, HARUO IKEDA
Qy	2303	uTyrAsnLleSerGlnProGlyThrGluLlePheAsnMetProAlaLleThrGlyThr--	2322		OS Streptomyces avermitilis
Db	19582	GTCTCTCC-AGCAAGTCGTGCGTCTTGTCTCGAGGACCTTCCCGCAGCAGGAACTGA	19640		PN WO 0162939-A/1
Qy	2323	-----GlyLeuMetThrTy 2327			PD 30-AUG-2001
Db	19641	GGAGCGCGGATGACGTTGAGCAGCGGAGACACAACCTTCGGCACTCTCATGGG	19700		PF 23-FEB-2001 WO 2001JP001381
Qy	2327	rArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaLleLeAr	2347		PR 24-FEB-2000 JP OOP 47405
Db	19701	TCGTCACGTTCCGAGCAGGAGAGAGCTG-----CTCAGCTCTGTCGCACTCCACTC	19754		PI HIROYUKI ENDO, HIROYUKI YAMAGUCHI, YUTAKA KANDA, SHINICHI
Qy	2347	gLysAlaLeuMetGlyLysTyrAspGlnTrpGluLysSerProLeuSerAlaAsnAl	2367		PI HASHIMOTO,
Db	19755	CGCGGACGTGCTCGGCGC---GACGACTCCGAGGCGCATCCCGCCGG-TCGGCTGTCA	19810		PI SATOSHI OMURA, HARUO IKEDA
Qy	2367	aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProLleThrAlaAlaS	2387		PC C12N15/54, C12N15/53, C12N9/10, C12N9/14, C12N1/21, C07C327/30, PC
Db	19811	GGGATCTAGGTTTCGACTCGCTTCGCGCGGTGGAGCTTCCCAACCACTCGCAGCAGA	19870		C12P17/08
Qy	2387	pGly-----ArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLy	2404		CC A method for procuding avermectin derivatives FH Key
Db	19871	CGGAGCTGGCTCTCGCGAGCTCTCGTCTTCGATTACCCAGCCCGCAGCAGCTGCC	19930		FT CDS (1)..(11916)
Qy	2404	sValSerGlyArgProSerSerArgLysAlaLysSerProAla-----	2418		FT CDS (11971)..(30687).
Db	19931	AATTCTGCTT-CCGAGATCGCGG---AGTTCAGCCCGCAGCACTCAACTCCGCTTCG	19986		Location/Qualifiers
Qy	2419	ProGlyLeuAlaSerGlyAspArgProProSerValSer---SerValHisSerG1	2436		1..30690
Db	19987	CGACCCCGGAGAGCTCGATGAGCCGATCGCCATCGTTGGCATGGCTGTGCTTCC	20046		/organism="Streptomyces avermitilis"
Qy	2436	uGlyAspCysAsnArgArgThr-----ProLeuThrAsnArgVa 2449			/mol_type="genomic DNA"
Db	20047	GGCGAGTGACCTCGCGGAGCACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATC	20106		/db_xref="taxon:33903"
Qy	2449	lTrpGluAspArgProSerSerAlaGly-----SerThrProPheProTy 2464			Alignment Scores:
Db	20107	GGCGGATTCCTCCCGACCGCGGCTGGACCTGACAGCTCTACGACGCGGCACTTCGAC	20166		Pred. No.: 1.59e-05 Length: 30690
Qy	2464	rAsnProLeu-----	2468		Score: 685.00 Matches: 704
Db	20167	CACCCGGCACTGCTACACCCGAAACGGCGGATTCCTTACGACGCGGCACTTCGAC	20226		Percent Similarity: 30.18% Conservative: 300
Qy	2468	eMetArgLeuGlnAlaGlyValMetAlaSerProProPro-----ProGlyLeuProAl	2486		Best Local Similarity: 21.16% Mismatches: 1193
Db	20227	GCCGAATTCCTCGGCATCAGCCCGCGGAAAGCCCTCGCATGGACCCCGACGAACTC	20286		Query Match: 5.18% Indels: 1144
Qy	2486	aglySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluProLysPro 2504			DB: 148
Db	20287	CTCTCGAAACCGCTCGGAAACCATCGAACCGCGGCATCAACCCCGCACCC 20341			US-09-522-753-5 (1-2517) x BD097649 (1-30690)
LOCUS	BD097649	30690 bp	DNA	linear	Qy 44 HisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArg 63
DEFINITION	A method for procuding avermectin derivatives.				Db 11193 CATCATCCGTTCCGCTGTCATCCCATGTCACCCGGA-----CGC 11234
ACCESSION	BD097649				Qy 64 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHis 83
VERSION	BD097649.1	GI:22643223			Db 11235 ACTGGCCCTCTTCGATGCGGC-----CCTGGCTCTGGACCGCGCTCTGCTGCC 11285
KEYWORDS	WO 0162939-A/1.				Qy 84 LeuArgProGluSerHisSerTyr-LeuProGluLeuGlyLysSerGluMetGluPhe1 103
SOURCE	Streptomyces avermitilis				Db 11286 CGCGAGCTGCTGCTCCGCCCGCCCGCTGCGCCCTGCGCCCTG-----CTGCAGGACCTCT 11336
					Qy 103 eGluSerLysArgProArgLeuLeuLeu-Leu-ProAspProLeuLeuArg---ProSerP 122
					Db 11337 GCCCGCACCGCGCGCGCACCAACCGCGCACCTACCGTGGTGGCGCAACCGCGC 11396
					Qy 122 roLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeu 142
					Db 11397 CCAGTGCAGCCCGCTGGCGCGCGAGACACAGAACACACACACACCTCTCTCGC 11456
					Qy 142 hrGlyLysLeuGluProValSerProProSerProHisThrAspProGluLeuGlu 162
					Db 11457 CTGGTCC---GCTCCACATCGCCACCGCTCTG-GGCCACACACACCC---GACA 11506
					Qy 162 euValProProArgLeuSerLysGluLeuLeuLeuGlnAsnMetAspArgValAspArg 182
					Db 11507 CCATCCCGCC-----GACCGCGCTTCGCG 11533
					Qy 182 luileThrMetValGlnGlnGlnIleSerLysLeuLysLysGlnGlnGlnGlu-Leu-Glu 201
					Db 11534 ACTTCGGCTTCGACTCCTCCTCAGCCGCTCGAATACGCAACCGCTCTCCCGCACCCCG 11593

RESULT 77

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Db	13576	GTGGGTACACCTCGCCACGCGCGCGGTTCGACACCGCGGCCACCTCATCGCC	13635	Db	14617	GACAACTCCCAACACCCGCCACACCACTCATCCCTCAACCCACCCACCCACCC	14676
Qy	751	ePro-----SerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr	769	Qy	1063	eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyAlaProProGly--H	1082
Db	13636	GCGGACGCGACACCTTCCTGCAAGCAGCTCCAGGCACTCGCGCAGCGCAACCCACCC	13695	Db	14677	CAAAACCACTCTCCCAACACCTCGCCAAACACCAACCACTGGCAGCCACCACTAC	14736
Qy	769	oProAlaThrLeuGlyAlaAsp-----GlyProProGlyProPro-----	783	Qy	1082	isProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProThrI	1102
Db	13696	GCGGTATCCACAGCAGCGCCGAGGGGAGCGGAGCGCGGAGCGCGAGAAAGACC	13755	Db	14737	ACCAACCA-----CAACCAACCCACACCCACCACTCGACCTCCACCC	14787
Qy	784	-----ThrProProArgArgThrSerArgAlaProLleGluProThr-----	797	Qy	1102	leSerAsnPro---ProProLeuLeuSer-----SerAlaLysHisProServ	1117
Db	13756	GCATTATCTGCTCCGACAGCGGACCCCAACGCGCGCGATGCCACCGGCTCTACAC	13815	Db	14788	TACCCCTTCAACACCACTACTGGCTCCAAACCCCGGCAAGCGGAGGACCC--G	14844
Qy	798	-----ProAlaSerGluAlaThrGlyAlaPro-----	806	Qy	1117	alLeuGluArgGln-----	1122
Db	13816	ACCCACCCGCTTCGCGCGGCGACTCAAGCATCTGACCCACCTCGACCCCCACCTC	13875	Db	14845	TCACGAGCGAAGCGCGTGCAGCAGCAGCAGCCCATCAACCCCGCTCGTGATGCTC	14904
Qy	807	-ThrProProAlaProProSerProSerAlaProProValProValProLysGluI	826	Qy	1122	leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTySerGlu-HisAla	1141
Db	13876	GACCAACCCCTCTCCCTCCCTCAGCAGGACCCCAACACCAAGGACACCAACCCCTC	13935	Db	14905	GTGGGCAAGTCTCCGACGAGCGAGACGAGAGCTGTTCGCGCTGGTGCACCCATCG	14964
Qy	826	uLysGluGluGluThrAlaAlaProProValGluGluGluGluGlnLysProPr	846	Qy	1142	LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu	1161
Db	13936	GAAGAAGCGCGCGCTGCTCCAGCAGACCGGCTACGCCAGCGCGGCTTCGCGCTTC	13995	Db	14965	GCGGCTGTCTGGGCGCATGCCCATCCCGAAGTGATGTTCCGAACAAGCGCTTCAAGAG	15024
Qy	846	oAlaAlaGluGluLeuAlaValAsp---ThrGlyLysAlaGluGluProValLysSerGl	865	Qy	1162	AlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro---	1180
Db	13996	CAGTTCGCTCCACCGCTCTCAGCAGCGCTACCATCATCAGCCCGCCCTACTACGCC	14055	Db	15025	CTGGGTTTGTATCTCTCGCCCAATTGATTCGATTCGATTCGATTCGATTCGATTCG	15084
Qy	865	uCyThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaGluAl	885	Qy	1181	-----ProGluSerLeu-----GlyValProThrAlaGlnGlu---AlaSerVal	1194
Db	14056	GGACACTCTCTCGCGGAATCAGCGCGCCACCTCGCGCATCTCTCAACCCCTCACCG	14115	Db	15085	CTGCGCTTCGCGCACCGCTGATTCGATTACCCCACTCCGATGCGCGCTTTCAGTTC	15144
Qy	885	aThrAlaGluGlyAlaLeuLysAlaGluLysLys-----	896	Qy	1195	LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSer	1214
Db	14116	GCCACACCTCATACCCAGCGCGCCACCTCATGCAACCAACCTGCGCGCGCCACCATG	14175	Db	15145	CTCGGCGCGCATGTCGCGAGCGACACAGC-----ACGACCACTCTGTCGCGCTA	15198
Qy	897	-GluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnLysSerAs	916	Qy	1215	ThrArgValProSerAspSerAlaIleThr-----TyrArgGly	1227
Db	14176	ACCACCTCCACACCCACCCACCAACATCACCACCACTCAGCGCGCGCAACAGAC	14235	Db	15199	ACTGCGTCCCGCGCGACGAGCGATCGCATCGTGGCATGCGCTGTGCTACCGCGT	15258
Qy	916	pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsuAr	936	Qy	1228	SerIleThr-----	1230
Db	14236	CTGCGCATGCGCGCATCAACACCCCACTCTCTGTCGTCAGCGCA-----	14284	Db	15259	GATGTACGAGCGGTGATGATCTTGGAGGTGTCAGTGGTGCCATGACGGATCGGC	15318
Qy	936	gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSe	956	Qy	1230	-----	1230
Db	14285	-----CCCCCACACGCTCAACATCATCA-----CCCTCTGCAACAA	14328	Db	15319	GGATTCGCGACGACCGTGGTGGACCTCGACACGCTGTACAAACCGGACCGGACAC	15378
Qy	956	rProGlnLysProLeuAspLeuLysGlnLysGlnArgAlaAlaIlePro--Pr	975	Qy	1231	HisGlyThr-----	1233
Db	14329	GGCATCAAAACCAAAACCTCCCAACCAACCAACCGCTTCCTCCATCCCGCCACCA	14388	Db	15379	CACGGAACAGCTACACCGGAGCGGGATTCCTTTACGACGAGGCAATTCGATGCC	15438
Qy	975	oIleGlnValThrLysValHisGluPro---ProArgGluAspAlaAlaProThrLysPr	994	Qy	1234	-----ProAlaAspValLeu-----	1238
Db	14389	ATCCTCAACCACTCCAGCAGACACCCCAACCTCCTCAGCCCGCCACCTACCGCC	14448	Db	15439	GACTTCTTGGTATAGTCCGCTGAGGCACTGGCATGAGCCCGGACGCGGCTGCTG	15498
Qy	994	oAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGl	1014	Qy	1239	-----TyrLysGly	1241
Db	14449	CTCATCAGCGCAACACCCCGGACCACTCTCAGCCCGCCACTACTGGACCAACAA	14508	Db	15499	CTGGAACACGCTGGGAGACATCGAACACGCTGATCAACCCCGACAGCTCCGTCGC	15558
Qy	1014	nProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAl	1034	Qy	1242	ThrIleThrArgIlelleGly-----GluAspSerProSerArg-----	1254
Db	14509	GCGCGCAACA-----CCGTGACTAGCGCACCAACCAACCCCTCCACCA	14556	Db	15559	ACCAACCGCGCTTCTTCGCGGGGTGACCTACACAGACTACCGCGCGGCTTCCACA	15618
Qy	1034	aPheAlaAlaGluAlaGlnLys-----LeuProGlyAspPro---ProCy	1048	Qy	1255	-----LeuAspArgGlyArgGluAspSerLeuProLysGly	1266
Db	14557	CACGGGCTCACCACTTACAGTCACTCGGACCGGACCAACCTCAGACCCCTCAGCC	14616	Db	15619	GCTCCGCGAGGTTTCAGGGGTATCTCGGACGGAAGCGCAGGAGCTATCGCTCGGT	15678
Qy	1048	strThrSerGlyLeuProPheProValPro-----ProArgGluValI	1063	Qy	1267	HisValleTyThrGlu---GlyLysLysGlyHisValleuSerTyThrGluGlyGlyMetSer	1285
				Qy	15679	CGTGTGCTACGCTCTCGGCTGGAAGGTGCGGCCCTCACAGTCGAC-----	15726

D	b	11634	ACCTCGGCTTGCATCCCTCACCGCGCTGGAACTTCAGCAACCGGCTCTCCGCCACCACCG	11693
Q	y	202	GluGluAlaLysProGluProGluLysProValSerPro-----Profile	219
D	b	11694	GACTCCGCTCCCACACCTCGCTTCGACACCCACCCACCACTCACCACCC	11753
Q	y	220	GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAenArg-LysLysAlaGl	239
D	b	11754	ACCTCCACACAACTCC-----AGCCAACAACCGGACACGCTGT--	11793
Q	y	239	wAlaAlaHiArgIleLeuGluGlyLeuGlyProGlnValGluLeu---ProLeuTyrAs	258
D	b	11794	-----CGCCCCGTGTGGGAGCTGCACAAACTCGAATCCGCCCTCTC-----	11838
Q	y	258	nGlnProSerAspThrArgGln-----TyrHi	267
D	b	11839	-CGCCCTCGAAAAACCGACAGCCAGCAAAGAGTGACCCCTCGGCTCAAGTCACTCA	11897
Q	y	267	sGluAenIleLysIleAenGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgAr	287
D	b	11898	TGTTGAGTGGGAAG-CACCCACGATCCGACAGCGAAAGCGCTGATCAGCAGAGAAG	11956
Q	y	287	gaEnHisalargaLysGlnTrpLysGlnLysPhe-----	298
D	b	11957	TTTCACT-----CGGCAACAGAGGCTGAGATTTCAAATTCAATTGACACGACTCGGC	12010
Q	y	299	CysGlnArgTyrAspGlnLeuMetGlnAlaLeuGluLysValGluArgIleGluAs	318
D	b	12011	CTGTCTGAACCGACCGCTGCCACTCCGCC-----CGTATCCGCTG	12052
Q	y	318	nAenProArgArg-----AlaLysGluSerLysValArgGluTyrTyrl	334
D	b	12053	-GGCCCTCTAGGAGTGAATGCAATTGGCGAATGAAGCAAGCTCTCGAATACCTCAA	12111
Q	y	334	uLysGlnPheProGluIleArgLysGlnArg-----GluLeuGlnGluArgMe	350
D	b	12112	GCGGCTCACTGCGGACCTGACCGCACTCGCCTCGCTGTACGAGGTGTGTCAGCGTGA	12171
Q	y	350	tGlnSerArgValGlyGlnArgGly-----SerGlyLeuSerMetSerAl	365
D	b	12172	GCAGGACCATCGGATTTGGGGATGGCGTGTCTTACCAGCGGGCGGAGTCAACC	12231
Q	y	365	aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGlu---	383
D	b	12232	CACGCGAGTGGCATCTCTCAAGTCCACAGCGGAGCTATCGGGAGTTCCCGACCGA	12291
Q	y	384	-----AsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTy	400
D	b	12292	CCGTGGATGAACCTGGAGCAG-----CTCTA	12318
Q	y	400	rAspAlaAspGlnGlnArg-----IleLysPheIleAenMetAsnGlyLeuMetAlaAs	418
D	b	12319	CGACCCGGACCCGACCGCTCAGAACCATGTTACACGCGAGCGGGAGGTTTCTATGA	12378
Q	y	418	p-----ProMetLysValTyrLysAspArgGlnValMetAsnMetTr	432
D	b	12379	CGCGGGCGACTTCGACCGCGCTTCTTCGAGTTGTCAACCGCTGAGCGCTGGCAATGA	12438
Q	y	432	pSerGluGlnGlu-----LysGluThrPheArgGluLysPheMetGln-----	446
D	b	12439	CCGCGACGCGCTGTCTGCTCGAAAACCACTTGGGAAACGTTCCGAGCGGGCGAATCGA	12498
Q	y	447	-----HisProLysAsnPh	451
D	b	12499	CCGAGGTCCATCGCGGAACCGGACCGGGGTTTTCTGTGGGATCAATCCGAGGACTA	12558
Q	y	451	e-----GlyLeuIleAlaSerPheLeuGluArg-----Ly	461
D	b	12559	CACCACCGGATACAC-ACATCAGCCCTCAACACGAGTTCGAGGGCTACCTGCTCACTGGCA	12617
Q	y	461	sThrValAlaGluCysValLeuTyrTyrTyrLeu-----ThrLysLysAsnGl	477
D	b	12618	GCGCGGCAAGCATTTGCTCAGGCGGTATCTCTTACCACTTCGGGCTCGAAGGCCCTGCGA	12677

QY	736	sSerGly-----ProAlaThrValAenAenSerSerAspThrGluSerI 751	Db	14717	GACAACTCCCAACACACCCCAACCACTCCCTCACCCACCCCAACCAACCC 14776
Db	13676	GTGGGTACACCTCCGCCACGCGCGCGGTTCGACACCGCGCCACCTCATCGCC 13735	QY	1063	eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyAlaProProGly---H 1082
QY	751	ePro-----SerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769	Db	14777	CAAAACCACTCTCTCAACAACTTCGCAAAACCAACCAACCACTGGGACCCCACTAC 14836
Db	13736	GCCGACCGGACACCTTCCTGCAAGACACTCCAGGCACTCCGCGCACTCCGCGGCAACCCCAACCC 13795	QY	1082	isProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProAspProProThrI 1102
QY	769	oProAlaThrLeuGlyAlaAsp-----GlyProProProGlyProPro----- 783	Db	14837	ACCACCAACCA-----CAACCAACCAACCAACCAACCAACCACTCCGACCTCCCAAC 14887
Db	13796	GCGGTATCATCACAGCAGCGCCCGAGCGGACCGGACCGGAGCGGAGCGCCAGGAAGACCC 13855	QY	1102	leSerAsnPro---ProProLeuIleSer-----SerAlaLysHisProSerV 1117
QY	784	-----ThrProProArgThrSerArgAlaProIleGluProThr----- 797	Db	14888	TACCCCTTCACACCAACCACTACTGGCTCCAAACCCGCAAGCGGAGCGACCC---G 14944
Db	13856	GCATTATCTGCTCCGACAGGCGACCCCAACGCGCGCGGATGCGCCACCGCTCTACCA 13915	QY	1117	alLeuGluArgGln----- 1122
QY	798	-----ProAlaSerGluAlaThrGlyAlaPro----- 806	Db	14945	TCACGAGCGAAGCGCGTAGCAAGCCACGACCCCATCAACCCCGCTGGGTGATGTCCTC 15004
Db	13916	ACCCACCCCGCTTCGCGCGCGCACTCAACGACATCTGCACCCACCTCGACCCCACTC 13975	QY	1122	leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyTySerGlu-HisAla 1141
QY	807	-ThrProProAlaProProSerProSerAlaProProProValProValProLysGluI 826	Db	15005	GTGGCAAGTCTCCGACGAGCGAGACGAGAGCTGTGGCGCTGTGGCGCACCATGCG 15064
Db	13976	GACCACCCCTCTCCCTCTCTCCAGGACCGGACCCCAACACCCAGGACACCAACCCCTC 14035	QY	1142	LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLeu 1161
QY	826	uLysGluGluGluThrAlaAlaProProValGluGluGluGluGluGluGluGluGluGlu 846	Db	15065	GCGGTGTGTGGCGCATGCGCACTCCCGAAGTGTTCGGAACAAGCGCTTCAAAGAG 15124
Db	14036	GAAGAAGCGCGCGCACTCTCCAGCAGACCGCGTACGCCGAGCGCGCTTCCTTCGCTTC 14095	QY	1162	AlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro--- 1180
QY	846	oAlaAlaGluGluLeuAlaValAsp---ThrGlyLysAlaGluGluProValLysSerG 865	Db	15125	CTGGGTTCGATCTCTCCGCGCAATTCAGCTTCGTATACCTGCTGCTGCTGCTGCTG 15184
Db	14096	CAGGTGCGCTCCACCGCTCTCTCCAGCAGCGGTACCACTACCCCGCTACTAGCC 14155	QY	1181	ProGluSerLeu-----GlyValProThrAlaGlnGlu---AlaSerVal 1194
QY	865	uCyThrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaGluAl 885	Db	15185	CTGGCGCTTCGCGCACCGCTGTATCTTCGATTAACCACTCCGATGGCGCTTTGGCAGTTC 15244
Db	14156	GGACACTCTCTCGGGAATCACCGCGCGCCACCTCGCGGCATCTCTCACCTCACCGAC 14215	QY	1195	LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyLysProSer 1214
QY	885	aThrAlaGluGlyAlaLeuLysAlaGluLysLys----- 896	Db	15245	CTCGGCGCGCGATCGTCGGAGCGGACACAGGC-----ACGACCACTGCTGCGGCTA 15298
Db	14216	GCCACCACTCATCACCAAGCGCGCACTCATGCAAAACCATGCGCCCGCGCACCATG 14275	QY	1215	ThrArgValProSerAspSerAlaIleThr-----TyrArgGly 1227
QY	897	-GluGlyGlySerGlyArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerA 916	Db	15299	ACTGGCGTCCCGCGCACGAGCGGATCGCATCGTCGCGATCGCTGCTGCTGCTGCTGCT 15358
Db	14276	ACCACCTCCACACACCCCGGACCATCACCATCACCATCACCATCACCATCACCATCAC 14335	QY	1228	SerIleThr----- 1230
QY	916	pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyLysLysAsnAr 936	Db	15359	GATGTACGAGCGGTGATGATCTCTGGCAGGTGTAGTGTGGGTGGCATACCGGATCGGC 15418
Db	14336	CTCGCATCGCGCATCAACACCCCGCACTCCCTCGTCTGTCAGCGCA----- 14384	QY	1230	----- 1230
QY	936	gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSe 956	Db	15419	GGATTCGCGACGAAACCGTGGTGGGACCTCGACACGCTGTACAAACCGGACCGGACCC 15478
Db	14385	-----CCCCCACACCGCTCAACACATCACCA-----CCCTCTGCGCAACAA 14428	QY	1231	HisGlyThr----- 1233
QY	956	rProGlnLysProLeuAspLeuLysGlnArgAlaAlaAlaIlePro---Pr 975	Db	15479	CACGGAACGAGTACACCGGAGCGGGATTCCTTTACGACGACGCAATTCGATGCC 15538
Db	14429	GGCATCAAAACCAAAACCTCCCAACCAACCGCTTCCTCACTCCCGCACCAACCC 14488	QY	1234	-----ProAlaAspValLeu----- 1238
QY	975	oIleGlnValThrLysValHisGluPro---ProArgGluAspAlaAlaProThrLysPr 994	Db	15539	GACTTCTTGGTATACGTCCGCTGAGGCACTGGCGATGGACCGGACGCGGCTGCTG 15598
Db	14489	ATCCTCAACCACTCCACGACGACACCCCAACCCCTCACCTACCCCGCCCGCACCC 14548	QY	1239	-----TyrLysGly 1241
QY	994	oAlaProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGlnG 1014	Db	15599	CTGGAACACGGTGGGAGAGCATCGAACGCGCTGCATCAACCCCGACAGCTCCGTCGC 15658
Db	14549	CTCATACCCGCAACACCCCGACCACTCTCAACCCCGCACTACTGAGACCAACAA 14608	QY	1242	ThrIleThrArgIleIleGly-----GluAspSerProSerArg----- 1254
QY	1014	nProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysGluAl 1034	Db	15659	ACACCAACCGGCTTCCTCGCGGCTGACCTACACGACTACCGCGGCGCTTCCACCA 15718
Db	14609	GCCGCAACA-----CCGTCGACTACGCGCACCAACCAACCAACCAACCAACCAACCA 14656	QY	1255	-----LeuAspArgGlyArgGluAspSerLeuProLysGly 1266
QY	1034	aPheAlaAlaGluAlaGlnLys-----LeuProGlyAspPro---ProCy 1048	Db	15719	GCTCGGACGGTTCGAGGGGTATCTCGGCGACGGAAGCGAGCAGTATCGCTCGCTCGGT 15778
Db	14657	CACGGCTCTACCACTCATCGAACTCGGACCGGACCAACCTCTACCACTCTACCACT 14716	QY	1267	HisValIleThrGlu---GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
QY	1048	sTrpThrSerGlyLeuProPheProValPro-----ProArgGluValI 1063			

Db 15779 CGTCTCGCTACGCTCTCGGCTCGAAGGTCCGGCCCTCACAGTCGAC----- 15826
Qy 1286 ValThrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAla 1305
Db 15827 ---ACTGCTGCTCTCTGCTCCT--GTCGCTCTGCA-----CTGGCCTGTACGGGCT 15876
Qy 1306 AlaProLysArgThrTyr-----AspMetMetGluGly 1316
Db 15877 GCGGTCCGGCGAGTGTTCATAGGCTCTCGGCTCGGGTGGGTGTACGGGTGATGTCAACCCCGGC 15936
Qy 1317 ArgValGlyArgAlaIleSerSerAla-----SerIleGluGlyLeu 1330
Db 15937 CGGGTTCGTGAGTTCGCGGCGAGCGGGGCTCGGCGGTGACGCGGGTGCAGAGGCTT 15996
Qy 1331 MetGly----- 1332
Db 15997 CTCGGCAGCGCTGACGGCACCGGCTGGGGTGGGGTGTGCGAATGCTGTGTTGGAGCG 16056
Qy 1332 ----- 1332
Db 16057 GCTGTGCGACGCGCGGCTCGGTCTCGGTACCGAATCTCTCGGTGTGTGGTGGCAGTGGCT 16116
Qy 1333 -----ArgAlaIlePro----- 1336
Db 16117 CAATCAGACGCTCGAGCAACAGGCTGACGGCGCCCAACGGCGCTGCCAGAGCGTGT 16176
Qy 1337 -----ProGluArgHis-----Ser 1341
Db 16177 CATCGGCTGGCCCTGCGCAAGCGGACCTGACCCCGCGCAGCTCGATCGGTGGAGGC 16236
Qy 1342 ProHisLysLeuLysGluGlnHisIle---ArgLysIleThrGlnGlyPro 1360
Db 16237 CCACGGCAC-----CGGCACCACTTTGGGCGACCGCATCGAGGCCCGGCTCTCT 16287
Qy 1361 ArgSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaLysLeuLeuLysArg 1380
Db 16288 CGCCACCTA-----CGG 16299
Qy 1381 GluGlyThrProProProProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db 16300 ACAGGCGGCGCGCAACGACCGCTGTGGCTCGATGAGTTCGACATCGGCCA 16359
Qy 1401 AlaLeuGlyProLysLeuLysProAlaHisGlu----- 1412
Db 16360 CGCGCAGGCTCCGCGAGTGTGGCGGGTTCATCAAGATGTGTGTCGCGAATGG 16419
Qy 1413 -----GlyLeuVal----- 1415
Db 16420 TCTGCTCCGCGGACGTTGATGTGTGATGACCGTCGCGCATGTGGACTGTCCGCGG 16479
Qy 1416 -----AlaThrValLysGluAla 1421
Db 16480 GCGGTGACGTGTCAGGAGAGCGTGCCCTGCGCGCGGGAGGGGGTGGCGG 16539
Qy 1422 GlyArgSerIle-----HisGluIleProArgGluGluLeuArgHis 1435
Db 16540 GGCAGGAGTGTATCTGTCGCGTTCAGCGGCACCAACGCCACGCTCATCTCGAAGAAGC 16599
Qy 1436 Thr---ProGluLeuProLeu----- 1441
Db 16600 ACCGCCCAACAATCCGTCAGACACACCGCGCAGACGCCCGCGGAGAGCGCGC 16659
Qy 1442 -----AlaProArgProLysLysGluLysIleThr 1452
Db 16660 CGAGGATGTTCCGGGGAAGCGCGCGGAGCAGCGCGGTACCGCGGGGAGC----- 16713
Qy 1453 GlnGlyThrProLysLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAsp 1472
Db 16714 GACTGGTCTGCTCCGCGCAGTCCAGGGGTGTGCGGTGTGTGTCGCGCAAGTCGCA 16773
Qy 1473 ValArgSerLeuIleGlySerProGlyArgThrPhePro----- 1486
Db 16774 GCGGCGCTGCGGCGCCAGGCGCCAGGCGCTGACAGCCCACTTCACCGACCAACCCCGGCT 16833

Qy 1487 -----ValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu 1500
Db 16834 CGACCTCGCCGACGCTCGGGTACACCT-----CGCCACGCGCG 16872
Qy 1501 ArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSer----- 1518
Db 16873 CGCGG-TGTTTCGACCAACCGCGCCACCTTCATCGCGCGCGACGCGACCTTCCTGCAAG 16931
Qy 1519 -----GlyGlySerIleAlaArgGlyAlaProVal 1528
Db 16932 CACTCCAGGCACCTCGCGCAGGCGAACCACCCCGCGCTCATCCACAGCAGCGCCCGAG 16991
Qy 1529 IleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla 1548
Db 16992 GCGGACCGGAGCGGGGAGCGCGAGAAAGACCGCATTCATCTGCTCCGACAGGCA 17051
Qy 1549 ProPheAlaGlyHisLeuProArg----- 1556
Db 17052 CCAACCGCGCGCATGGCCCGCTTACCACACCCACCGCTCTTCCGCCCGCGCAC 17111
Qy 1557 -----GlySerProValThr 1561
Db 17112 TCACAGCATCTGACCCACCTCGACCCCGCCTCGACCCCTCTCTCCCTCTCTCA 17171
Qy 1562 MetArgGluProThrProArgLeu----- 1569
Db 17172 CCAGGACCCCAACACCCAGGACACACCCCTCGAAGAGCGCGCAGCTGCTCCAGC 17231
Qy 1570 -----GlnGluGlySerLeuSerSerSerLysAlaSerGln 1581
Db 17232 AGACCCCGTACGCGCGCGCTCTTTCGCTTTCAGGTTCGCGCTCCACCGCTCTCTCA 17291
Qy 1582 AspArgLysLeuThrSerThrProArgGluIle-----AlaLysSerPro 1596
Db 17292 CCGAGGCTACCATCACCCCGCTACTACGCGGAGACTCTCTCGGGAATACCG 17351
Qy 1597 -----HisSerThrValPro-GluHisProHisProIleSerPr 1610
Db 17352 CGGCCACCTCGCGGCTCTCTCCCTCACCGGCGCCACCGCTCATCACCAACGCG 17411
Qy 1610 oTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLe 1630
Db 17412 CCACCTCTCGAACAACCTCGCGCGCACCATGACACCGCTCCACACCGCGCCAC 17471
Qy 1630 uAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAlaIleTyr-- 1649
Db 17472 ACATACCCACCTCACCGCCACGAAACGACCTCGCGCATGCGCGCATCAACACCC 17531
Qy 1650 -TyrLeuProArgHisLeuAla----- 1656
Db 17532 CCACCTCTCTGTCATCAGCGCGCACCCCGCACCGTCCACACATCACCGCTCTGCC 17591
Qy 1656 ----- 1656
Db 17592 AACACAAGGCATCAAAACCAAAACCTCCCGCACCAAAAACGCTTCACCTCCCGCCACA 17651
Qy 1657 -----ProAsnProThr-----TyrProHisLeu-TyrProProT 1668
Db 17652 CCAACCCCATCTCAACCAACTCCACGAGCACCCCAACCCCTACCTACCGCCACCC 17711
Qy 1668 yr-----LeuIleArgGlyTyrProAspThrAlaAlaLeu----- 1679
Db 17712 ACACCCCTCATCAGCGCAACACCCCGCGACCACTCTCCCGCGCTACTTGA 17771
Qy 1680 -----GluAsnArgGlnThrIleLeuAsnAspTyrIleThrSerGlnMethHisHis- 1697
Db 17772 CCAACAAGCGCGCAACACCGTC-----GACTACGCGCACCAACCGCAACCGCTCCAC 17825
Qy 1698 AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlu 1717
Db 17826 AACAGCGCGTCACCACTACATCGAATCGAGC-----CGGCAACA 17867

QY 1718 SerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleLeuAspLeuSerGlnVal 1737
 Db 17868 CCTCA----- 17873
 QY 1738 ProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaMetAspArg 1757
 Db 17874 CCACCTTACCCACCAACACTCCCAACACACCCCAACCACTCCACCTTCACTCCACCCAC 17933
 QY 1758 LeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerProLeu 1777
 Db 17934 CCCACC-----ACCACCCCAACACCACTCTCTCCCACTCGGCCAAACACCA 17984
 QY 1778 SerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGlu 1797
 Db 17985 CCACCTGGCACCCACC-----ACTACACCCACCAACCAACCAACCC-CACACCCAC 18037
 QY 1798 ArgAspArgAspArgGluArgAspArgAspArgGluArgGlyLysSerIleLeuThrSer 1817
 Db 18038 ACCCACCTCGAC-----CTCCCC 18055
 QY 1818 ThrThrThrValGluHisAlaProIleTrp-----ArgProGlyThrGluGln 1833
 Db 18056 ACCTACCCCTTCCAAACACAGCACTACTGGCTCGAAGACACACAGCGGGTGGCGGATCC 18115
 QY 1834 SerSerGlySer-----SerGlySerSerGlyGlyGlyGlySerSer----- 1848
 Db 18116 GGTTCGGGTTCCGGTTCGGGGCGGAGGACTCGCGGGCGGAGCGGAGAGGTGGAGTGG 18175
 QY 1848 ----- 1848
 Db 18176 CGGTTCTGGGAGCGGTGGCCCGCAGACCTGGAAGCTGGGACCACTCGCCGTG 18235
 QY 1849 -----SerArgProAlaSerHisSerHisAla 1857
 Db 18236 CCCCCCTCCGGCGCTGGACACGGTGGTGGCGCACTCTCCGGCTGGCACCGGCCACAA 18295
 QY 1858 HisGln----- 1859
 Db 18296 CACGACCAAGCCCGCATCAACACCTGGACCTGACGAGAACTGGAAACCCCTCACCCCTC 18355
 QY 1860 -----His-SerProIleSerProArgThrGlnAspAlaLeuGlnArgProSe 1876
 Db 18356 CCCACACCCCAACACCCCAACCACTGGCTCATCGCCATCCCGAAGACCAACCAAC 18415
 QY 1876 rValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPro-----SerLysPr 1895
 Db 18416 C-----ACCCCAACATCAACCAATCTCCACC 18442
 QY 1895 oThrValLeuArgSerThrSerSerSerProValArgProAlaAlaThrPheProPr 1915
 Db 18443 AAAC-----TCCACACACCGGCATCACCCCTCCCTCCCTTCACTCCACCAAC 18493
 QY 1915 oAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluPr 1935
 Db 18494 C-----ACACCAACCCCAACCTCCACCAACCCGACCAACCAACCAACCAAC 18550
 QY 1935 oValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspTh 1955
 Db 18551 GGACCCATCACCGGCTGCTCTCCCTCGCCCTCGCCCTCGAGAAACCAACCCCAACCC 18610
 QY 1955 rGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerPr 1975
 Db 18611 CCCC-----ACACACCCACCGGCACCTCTCAACCTCACCTTCACTCCCAACCC 18658
 QY 1975 oSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAl 1995
 Db 18659 CACA-----CCCAACCCCAACCCCAACCCCTCTGTGTGGCCACCAACCAAC 18706
 QY 1995 argThrProAlaLysAsnLeuAlaProHisAlaSerPro----- 2009
 Db 18707 GCCACCAACCAACCCCAACCCCTCTCACACACCCCAACCCCAACCCCAACCTGGGA 18766
 QY 2010 -----AspProProAlaProProAlaSerAlaSerAspPr 2021

Db 18767 CTCGCCCGCACCCCTCTCGAACACCCCAACCAACCCCGGATCATCGACCTC-CC 18825
 QY 2021 o-----HisArgGlyLysThrGlnSerLysProPheSe 2032
 Db 18826 CACCACCCCAACCCCAACCACTCCACCACTCCCAACCACTCCCAACCACTCCCAAC 18885
 QY 2032 rIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluG 2052
 Db 18886 CCAAAACCACTCGCCATCGCCACCAACCGGACCCCAACCCCGCGCTCACCCCAAC 18945
 QY 2052 yValGluProValSerProValSerSerProSer-----LeuThrHi 2066
 Db 18946 CTTACCCCAACCAACCAACCAACCCCAACCCCAACCCCAACCACTCATCGCG 19005
 QY 2066 sAspLysGlyLeuProLysHisLeuGluLeuAspLysSerHis----- 2081
 Db 19006 CGGAACCGCGG-CCTCGCCACCACTCACCAACCACTCACCAACCAACCAACCC 19064
 QY 2082 -----LeuGluGlyGluLeuArgProLysG 2090
 Db 19065 AACACCTCTCTCTACAGCGAAGCGGCCCAACCCCAACCCCAACCACTCACCA 19124
 QY 2090 nProGlyProValLysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuPr 2110
 Db 19125 CCNACTCCAAACAA-----AGGCATCCACTCACCACTCACCACTCGC 19169
 QY 2110 oGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisG 2130
 Db 19170 ACACGACGACCAACCACTTCA-----ACAACCTCT-----CAACACCA 19211
 QY 2130 nArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHi 2150
 Db 19212 TCCCCCAACCAACCCCTCACCA-----CGTCATCCACCGCAGGCA 19256
 QY 2150 s-----HisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPh 2166
 Db 19257 TCTCGACGACGCACTCCCTCAAC-----CTCACCCCAACCACTCAACCAAG 19307
 QY 2166 eProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyrLeuPr 2186
 Db 19308 TCTCGCGCCCAAGCCCA-----CAGGCCCACT-----CC 19340
 QY 2186 oProProAspHisGlyAla-----ProAlaArgLysPro----- 2198
 Db 19341 TCCACCACTCACCAACCAACCACTCCCTCAACGCTCTCTCTCTCTCTCTCTCTCTCT 19400
 QY 2199 -----HisSerGluGlyGlyLysArgSerPr 2207
 Db 19401 CCACCTTCGGCGCACCCCGCCCAAGCACTAGCCCGCAGCAACCGCTACCTCGAGGCC 19460
 QY 2207 oGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerPr 2227
 Db 19461 TCGCCCA----- 19469
 QY 2227 oProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrAr 2247
 Db 19470 ACCGCCACACCCCACTCCCGCCCAAC-----CATCG 19505
 QY 2247 aAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerG 2267
 Db 19506 CTTGGGCACTGGCAAGG---AAACGACTGGCGACTGGTCAAGTCAGCAACATCTCC 19562
 QY 2267 nProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLys 2287
 Db 19563 GCGCGCGCGGA-TGTTCCGATCGCGCGAGTGGCGGTGTCAGCTGTGTGCGCGCG 19621
 QY 2287 sGlnGluIleAsnLysLysLeuAsnThrHis-----AsnArgAsnGluProG 2303
 Db 19622 ATCCGACGCGCGCGCGAGTCTCTCTCGCCGATATCCATCGGAAGAAATTTGGGACCG 19681
 QY 2303 uTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThr-- 2322

Db 19682 GTTCTCTCC-AGCAAGTCGTGGTCTTGTCTGAGGACCTTCCCGCAGCAGGAACTGA 19740
Qy 2323 -----GlyLeuMetThrTy 2327
Db 19741 GGAGCGCGGAGTACCTGTGAGCAGACGAGAGCACAACCTCCGGCACTCTCATCGG 19800
Qy 2327 rArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaLeuIleAr 2347
Db 19801 TCGGTACGTTCCGAGCAGGAAGAAGAGCTG-----CTCAGCCTCGTCGGCATCCACTC 19854
Qy 2347 gLyAlaLeuMetGlyLysTyArgGlnTrpGluLysSerProProLeuSerAlaAsnAl 2367
Db 19855 CGCGGAGTGTCTGGCGGC---GACGACTCGAGCCATCCCGCCCGG-TCGGCTGTTC 19910
Qy 2367 aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
Db 19911 GGGATCTAGGTTCCGACTCGTTCGGCGGTGGAGCTTCGCAACCACTCGCAGCAGAGA 19970
Qy 2387 pGly-----ArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLy 2404
Db 19971 CGGAGCTGGCTCTCGCAGCACTCTCGTCTTCGATTACCCAGCCAGCCAGCTCGCC 20030
Qy 2404 sValSerGlyArgProSerArgLysAlaLysSerProAla----- 2418
Db 20031 AATTTCGTCT-CCGAGATCGCG---AGTTCAGCCCGCAGCACTCAACTCCGCTTCG 20086
Qy 2419 ---ProGlyLeuAlaSerGlyAspArgProProSerValSer---SerValHisSerGl 2436
Db 20087 CGACCCCGGCGAGCTCGATGAGCCGATCGCCATCTGTGGCATGGCTTCGCTTCCC 20146
Qy 2436 uGlyAspCysAsnArgThr-----ProLeuThrAsnArgVa 2449
Db 20147 GCGGAGTACCTCGCGGAGCACTTCTGGATCTGATCTCTCCGAGCAGGAGCGCATC 20206
Qy 2449 lTrpGluAspArgProSerSerAlaGly-----SerThrProPheProTy 2464
Db 20207 GCGGATTCCTCCACCGCGCGGTGGACCTGGACACGCTCTACGACCCCGACCC 20266
Qy 2464 rAsnProLeu----- 2468
Db 20267 CACCCGCGCCTGTCTACACCCGAGCGGGATCTCTTACGAGCGGCGCCTTCGAC 20326
Qy 2468 eMetArgLeuGlnAlaGlyValMetAlaSerProProPro-----ProGlyLeuProAl 2486
Db 20327 GCCGAATCTTCGGCATCAGCCCGCGGAGAGCCCTCGCATGAGCCCGCAGCAAGACTC 20386
Qy 2486 aGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluProLysPro 2504
Db 20387 CTCTCGAACCCTCGGAAACCATCGAAGCGCGCATCAACCCCGACCC 20441

RESULT 79
LOCUS AP005024 302675 bp DNA linear BCT 10-MAY-2003
DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section 4/30.
ACCESSION AP005024 BA000030
VERSION AP005024.1 GI:29604389
KEYWORDS Streptomyces avermitilis MA-4680
SOURCE Streptomyces avermitilis MA-4680
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
TITLE Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403
PUBMED 11572948
REFERENCE 2

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
CDS
CDS
CDS

Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.
Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
Nat. Biotechnol. 21 (5), 526-531 (2003)
22608306
12692562
3 (bases 1 to 302675)
Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kishida,N., Director-General of Biotechnology Center, Shiba,T., Sakaki,Y. and Hattori,M.
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:bioinfo.go.jp, URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)
This work was done in collaboration with Haruo Ikeda(*1), Jun Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Osonoe(*4), Norihiro Kishida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3).
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.ls.kitasato-u.ac.jp.
Location/Qualifiers
1. 302675
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227882"
/notes="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"
12. .425
/note="SAV750"
/codon_start=1
/transl_table=11
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC68460.1"
/db_xref="GI:29604390"
/translation="MAELCLDREARTAAQAQSLGAGCTVARNLDAVALDFDSGG
PMSGLAMTRPAALRIREALVDGAGFTQAIEATGYLKAERGLGRVAETVDRPTVA
LAIVGTAHLMLTWSFGAPDRPHMARLVAALVDS"
1504. .2592
/note="SAV751"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68461.1"
/db_xref="GI:29604391"
/translation="MRTVGVEELLVDPETGPKALSTAVLARAEOVDPDQVFEKE
LHGQMLEFATHPTDMAALGAEIVRCRKEARHAGGCAVALATSPVPSIANN
RYYQWAEQYGTAMQQLTCGCHVHVAVSDSEGVAVDRIPLPVLVALSANSFW
QGRDSYYSRSRWGRWFSAGTFLPGSPERVHRQVADMTATGVLDDGVYFDARL
SHRYPTVRVADVCLHPDTAQIATLARGVESAARDMRAGREDDHGSVLLRLAAW
QARAGLSGELLHPVTMRIRAESVVRALLIEHVGLDALADGLDLAHEAVAEILLERN
GARVQRELLERTGLRDTVTECVROTQG"
complement(3011..3400)
/note="SAV752"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

/protein_id="BAC68462.1"
/db_xref="GI:29604392"
/translation="MOITASTVSTIVADVAASQAFITAHLYGTVQAAADGCFASLSTRAD
ADVILLARGTQVLPPQRDRHAGLLIATFLADGLHDQEKRLREAGVEITWPLREEP
WGERLFOITDPNGVIIQFVWATSDDT"
complement(3457..3648)
/note="SAV753"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68463.1"
/db_xref="GI:29604393"
/translation="MFELDRTTVSALTIGLGEVPAGRDVDPAPRGGYDLURLLAADRA
VRGRPRPRVRNIPRRRGY"
3755..4408
/note="SAV754"
/codon_start=1
/transl_table=11
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC68464.1"
/db_xref="GI:29604394"
/translation="MGTEITSPAQVPLGRSKKHETRLQLSDHATRLFLAQGFQOTTI
AEIAAARMAKTVTNPFAKEDLDHQBETASLANAVADRRPGSALAAILLRFAF
DSLATQNPVAGSCPAFABWADSPTLTACLRGLHDQREHRLAQALADADATDTPDDIT
VHIAAGLLGTVHRVLFQRIQDLTLAGHPNDRIAHTITAEATQAFDLLEPALAHAYATA"
4587..5627
/note="SAV755"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="BAC68465.1"
/db_xref="GI:29604395"
/translation="MISPKRIRTTTSPVACATGSPSQATSAFPAVTRACRSTWT
CRSDIRRPAPAPVPRSQSADFAPLVRSILDTGCRHVLRSVLLSDFDVP
LVTRAGTTHADPALRLTERFVFNENPQTCGAAPPAAGHPDRWAVSGRPKSDGP
APTSTPAASHPAGRNQHRGSAVVTGMEAVTPKEWLAVLAVSLDAVHTGREALRHLR
AARGNVSGAGLGDYGMWAAVNAKGAVNLTAMALDHAPEVRVNAHGPGFTTHSG
TAGATPDTPVRSFAERILPLGRVAEPDEIASVVAFLADPGDALYDRRSNPGRRRPHRC
YAAAPHRDVNAS"
complement(6101..8071)
/note="SAV756"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68466.1"
/db_xref="GI:29604396"
/translation="MERSKDRRPSVLQDRARYWEDTRLARSTGLMGWLVIICLAVVV
PVYALLVMTDPSRSLRGLAAVWRSAETRLGTVTGLRMLLSALLGLVLCV
STLGVITTTGERALAEISRGSTVLEQHVAGVSPASDFGVIALVSPQASTVLVLP
VVLAERDTERALAEISRGSTVLEQHVAGVSPASDFGVIALVSPQASTVLVLP
EPTADAEVLRLVLLALGEGTGGPVLLAVLDVRYAPARLAAGPRGTVLETDV
ARLIAQVCRGLSLVRLDLTDFAGDEHLEATAHGHPFGAALLGHATSCVVGLLT
AGERTLNPANTLVAGSRVLVLTDDGVPAPEDCRHLVPSATAMAQPPQSDALS
LILGWNRLAPLVNQLRTPARPGSVLDVTDRAVGPTRPEPSGAARLAVRFSADLS
RETLGLDLAPYDGLIIVLGPFGDGPDRDRTLVTLALRLLEERTGRENRYTEL
TQDRNRLPVPNGPSDVIYSGELTGLLQAIAQNRHLRAIPPDFELFSADGSTICLRQA
HYVITGCTTTFATVAAARDGECAIGYRHDRRAIPPDHGIRLNPDKSRERWSAD
QVVVATEPTEPAAGQDTBHQPGPADTRA"
9086..10072
/note="SAV757"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68467.1"
/db_xref="GI:29604397"
/translation="MGRRKGGIGIALVTSVMLVGASACGGSGDKANGDDAKASGKP
ASASPKPKAGPMLLETITPQTGTGVKAMPISVFTNPAAKARAARAEKHMVVS
SQPVSAHWFGDKRADMRPKTYWSPGTGKPIDADMKNVNGNGRYGVHYSVTHTKPK
DDVRAVSVYGHMKTRNGKVVRLTISINAGSAQPTWNGTMAVIDKQEKVHTMTCSD
GISCEKGSFNYYDLTLPDWDHLTQSGTYVITSGDPTPGSGARSCHGCVHLSMSDA
WFYGVQKQGDPPVITGSPRAKAPADNGYADFNGLMDQWILGASGEKKTATIL"
complement(11089..11494)
/note="SAV758"

```

```
Db 249350 ACCTCGGCTTCGACTCCCTCAACCGCGTTCGAACTACGCAACCGGCTCTCCCGCACCAACCG 249409
Qy 202 GluAlaAlaLysProGluProGluProGluProValSerProPro-----Proile 219
Db 249410 GACTCCGCTCCACACACCTCGCTTCGACACACCCACCCACCCCTCACCCACC 249469
Qy 220 GluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArg-LysLysAlaG1 239
Db 249470 ACCTCCACACAACTCC-----AGCCACAACCGGACAACTGT-- 249509
Qy 239 uAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeu---ProLeuTyrAs 258
Db 249510 -CGCCCGCTGTGTGGGAGCTCGACAACTCGAATCCGCGCTCTC----- 249554
Qy 258 nGlnProSerAspThrArgGln-----TyrHi 267
Db 249555 -CGCCCTCGCAAAACCGACGCGCCGAGGAGAGTCACTCCGCGCTGAAGTCACTCA 249613
Qy 267 sGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgAr 287
Db 249614 TGTAGGTGGAACG-CACCCAGCATCCGACACCGAAAGCGCTGATGACGACGAGAAG 249672
Qy 287 gAsnHisAlaArgLysGlnTrpLysGlnLysPhe----- 298
Db 249673 TTCACAT-----CGGCAACAGAGGCTGAGATTTTCAAATTCAATTGACAACGACTCGGC 249726
Qy 299 -CysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAs 318
Db 249727 CTGTCTGAACCGAGCGCTGCCACTCGCC-----CGTATCCGCTG 249768
Qy 318 nAsnProArgArgArg-----AlaLysGluSerLysValArgGluTyrTyrG1 334
Db 249769 -GGCCCTGCTAGGACGTGAATGCAATTGGCGAATGAGCGAGCTCTCGGAATACCTCAA 249827
Qy 334 uLysGlnPheProGluIleArgLysGlnArg-----GluLeuGlnGluArgMe 350
Db 249828 GCGCGTCACTCGGACCTGGACCGCACTCGCGCTGCTGTACGAGGTGTCGAGCGTGA 249887
Qy 350 tGlnSerArgValGlyGlnArgIly-----SerGlyLeuSerMetSerAl 365
Db 249888 GCAGGACCGATCCGATTTGGGGATGGCGTGTCTGTCATCCAGCGGGGCGAGCTCACC 249947
Qy 365 aAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGlu----- 383
Db 249948 CACGCGACTGGGCATCTCTGTCAGTCCAGACGACGCTATCGGGAGTTCCCGACCGA 250007
Qy 384 -----AsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeuTy 400
Db 250008 CCGTGGATGGAACCTGGAGCAG-----CTCTA 250034
Qy 400 rAspAlaAspGlnGlnArg-----IleLysPheIleAsnMetAsnGlyLeuMetAlaAs 418
Db 250035 CGACCGGACCCCGACCGCTCAGGAACAGATTACACCGCAGCGGAGGGTTCTCTATGA 250094
Qy 418 p-----ProMetLysValTyrLysAspArgGlnValMetAsnMetTr 432
Db 250095 CGCGGGGACTTCGACCGCGGTTCTTCGAGTTGTACCCGGTGAGCGCTGGCAATGA 250154
Qy 432 pSerGluGlnGlu-----LysGluThrPheArgGluLysPheMetGln 446
Db 250155 CCGCAGCAGCGCGCTGCTCGAAACCACTTGGGAAACGTTGAAACGCGCGGAATCGA 250214
Qy 447 -----HisProLysAsnPh 451
Db 250215 CCGAGGTCCATGCGCGGAACCGGACCGGGTTTTCTGGGGATCAATCCGAGGACTA 250274
Qy 451 e-----GlyLeuIleAlaSerPheLeuGluArg-----Ly 461
Db 250275 CACCACCGGATACAC-ACATCAGCCCTCAACCGCAGTTCGAGGGTACTCTACTGGCA 250333
Qy 461 sThrValAlaGluCysValLeuTyrTyrTyrLeu-----ThrLysLysAsnG1 477
```



```
QY 736 sSerGly-----ProAlaThrValAsnAsnSerSerAspThrGluSerIle 751
Db 251392 GTCGGGTACACCTCGCCAGCCCGCGCTGTTTCAGACACCGCGCCACCTCATCGCC 251451
QY 751 ePro-----SerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysPr 769
Db 251452 GCGACCGGACACCTTCTCGAAGCACTCCAGGCATCGCCGACGGCGNACCCACCC 251511
QY 769 oProAlaThrLeuGlyAlaAsp-----GlyProProGlyProPro----- 783
Db 251512 GCGGTATCATCAGCAGCGCCCGAGCGGACCGGAGCGCGGAGCGCGAGGAAAGACC 251571
QY 784 -----ThrProProArgThrSerArgAlaProIleGluProThr----- 797
Db 251572 GCATTTCATCTGCTCCGACAGGCGACCCAAACGCGCGGCGATGCCACCGCCTCTACAC 251631
QY 798 -----ProAlaSerGluAlaThrGlyAlaPro----- 806
Db 251632 ACCACCCGCTCTCGCGCGGCACTCAAGCATCTGCACCCACCTCGACCCCCACCTC 251691
QY 807 -ThrProProAlaProProSerProAlaProProProValProValProLysGluI 826
Db 251692 GACCACCCCTCTCTCCCTCTCACCAGGACCCCAACACCCAGGACACCCACCCCTC 251751
QY 826 uLysGluGluGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProPr 846
Db 251752 GAAGAAGCGCGCGCACTGCTCCAGCAGACCGCTACGCCCGCGCCCTCTTCGCGCTC 251811
QY 846 oAlaAlaGluGluLeuAlaValAsp---ThrGlyLysAlaGluGluProValLysSerGl 865
Db 251812 CAGTCCGCCCTCCACCGCTCTCCAGCGGCTACCATCATCACCCTCCACTACTAGCC 251871
QY 865 uCysThrGluGluAlaGluGluGlyProAlaLysGlyLysAlaGluAlaGluAl 885
Db 251872 GGACACTCCCTCGCGGAAATCAGCGCGCCACCTCGCGGCATCTCTCACCCTCACCGAC 251931
QY 885 aThrAlaGluGlyAlaLeuLysAlaGluLysLys----- 896
Db 251932 GCCACCACTTCATCCCAACCGCGCCACCTCATGCAACCATGCCCCCGGACCATG 251991
QY 897 -GluGlyGlySerGlyArgAlaThrAlaLysSerSerGlyAlaProGlnAspSerAs 916
Db 251992 ACCACCTCCACACCCCCACCATCATCACCACCTCACCGCCCAAGAAACGAC 252051
QY 916 pSerSerAlaThrCysSerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnAr 936
Db 252052 CTCGCCATCGCGGCATCAACACCCCACTCCCTCCGTCATCAGCGGCA----- 252100
QY 936 gLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSe 956
Db 252101 -----CCCCCCACACGTCACATACCA-----CCCTCTGCCAACACAA 252144
QY 956 rProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIlePro---Pr 975
Db 252145 GGCATCAAAACCAAAACCTCCCAACCAACACCGCTTCACCTCCCAACCAACCC 252204
QY 975 oIleGlnValThrLysValHisGluPro---ProArgGluAspAlaAlaProThrLysPr 994
Db 252205 ATCTCTCAACCACTCCACAGCACACCCAAACCTCTACCTACCAACCCCAACCC 252264
QY 994 oAlaProProAlaProProProGlnAsnLeuGlnProGluSerAspAlaProGlnGl 1014
Db 252265 CTCTATCAGGCAACACCCCAACCAACCACTCTCACCACCCCACTACTGACCCCAACAA 252324
QY 1014 nProGlySerSerProArgGlyLysSerArgSerProAlaProAlaAspLysGluAl 1034
Db 252325 GCGCGCAACA-----CGTCCGACTAGGCGCACCAACCAACCCCTCCACCA 252372
QY 1034 apheAlaAlaGluAlaGlnLys-----LeuProGlyAspPro---ProCys 1048
Db 252373 CACGCGGTACACCATCATCGAACTCGGACCCCGAACACACCTTCACCCCTTCACCCAC 252432
QY 1048 sTrpThrSerGlyLeuProPheProValPro-----ProArgGluValIle 1063
Db 252433 GACAACTCCCAACACACCCCAACCACTCACCTCACCCACCCCAACCAACCC 252492
QY 1063 eLysAlaSer-ProHisAlaProAspProSerAlaPheSerTyAlaProProGly---H 1082
Db 252493 CAAACCCACTCTCTCACCACCTCGCCAAACACCAACCACTCGGCACCCCACTAC 252552
QY 1082 iSProLeuProLeuGlyLeuHisAspThrAlaAspProValLeuProArgProProThrI 1102
Db 252553 ACCCAACCA-----CAACCAACCAACCAACCAACCACTCGACCTCCCAAC 252603
QY 1102 lSerAsnPro---ProProLeuIleSer-----SerAlaLysHisProSerV 1117
Db 252604 TACCCCTTCCACACCACTACTGCTTCCAAACCAACCGGCAAGCGAGCGACCC---G 252660
QY 1117 alLeuGluArgGln----- 1122
Db 252661 TCACGAGCGAAGCGGTGAGCAAGCCACGACCCCATCAACCCCGCTCGTGATGTCCTC 252720
QY 1122 leGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyTrSerGlu-HisAla 1141
Db 252721 GTGGGCAAGTCTCCGAGGAGGAGAGAGAGTGTTCGCGCTGTGGCGCACCATGCG 252780
QY 1142 LysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLysLeu 1161
Db 252781 GCCGTGTCTGGCATGCCACTCCGAAAGTATGTTCCGAAACAAAGCCCTTCAAAGAG 252840
QY 1162 AlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAlaGlyPro--- 1180
Db 252841 CTGGGTTTGTATCTCTCGCCCAATTCAGTTCTGTAATCGACTGCTTCTGACGTGAC 252900
QY 1181 -----ProGluSerLeu-----GlyValProThrAlaGlnGlu---AlaSerVal 1194
Db 252901 CTGCCCTTCCGCGCACGCTGATCTTCGATTACCCACTCCGATGGCGTGTTCGACGTTC 252960
QY 1195 LeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSer 1214
Db 252961 CTCGGGGCGCGATCGTCGGAGGACACAGGC-----ACGACCACTCGTCTGCGCTA 253014
QY 1215 ThrArgValProSerAspSerAlaIleThr-----TyrArgGly 1227
Db 253015 ACTCGCGTCCCGCCGACAGCGATCGGCATCGCCATCGTCGGCATCGCTGTCGTACCCCGT 253074
QY 1228 SerIleThr----- 1230
Db 253075 GATGTACGAGCGTCTGATCTCTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 253134
QY 1230 ----- 1230
Db 253135 GGATTCGCGAAGAACGCTGGTGGGACCTCGACACGCTGTACAAACCCGAGCCGACAC 253194
QY 1231 HisGlyThr----- 1233
Db 253195 CACGGAACACAGCTACACCCGAGCGGGGATTCCTTTACGACGAGGCAATTCGATCCC 253254
QY 1234 -----ProAlaAspValLeu----- 1238
Db 253255 GACTTCTCGTATCAGTCGCGTGGAGCAGCTGGCGATGGACCCGAGCAGCGGTGCTG 253314
QY 1239 -----TyrLysGly 1241
Db 253315 CTGGAACACGCTGGGAGAGCATCGAAACAGCCCTGTCATCAACCCGACAGCTCCGCTGGC 253374
QY 1242 ThrIleThrArgIleIleGly-----GluAspSerProSerArg----- 1254
Db 253375 ACACCAACCGCGCTTTCGCGCGGTGACCTTACCACTACGATACCGCGCGCTTCCCA 253434
QY 1255 -----LeuAspArgGlyArgGluAspSerLeuProLysGly 1266
Db 253435 GCTCCGCGCAGGGTTCGAGGGGTATCTCGGCGACGGAAGCGAGGAGTATCGCTCGGT 253494
QY 1267 HisValIleTyrGlu---GlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSer 1285
```



```
Db 257398 GTTCTCTCC-AGCAAGTCGTGGTCTTCTCGAGGACCTTCCCGCAGGACGAACTGA 257456
Qy 2323 -----GlyLeuMetThrTy 2327
Db 257457 GGAGCGCGCAGTACCGTTGAGCAGACGAGACACAACTCCGGCACTCTCATGGG 257516
Qy 2327 rArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleLeAr 2347
Db 257517 TCGGTACAGTTCGCGAGCAGGAAGAAGAGCTG-----CTCAGCCTCGTCCGCATCCACTC 257570
Qy 2347 gLyseAlaLeuMetGlyLysTyArgGlnTrpGluSerProProLeuSerAlaAsnAl 2367
Db 257571 CGCGGCAGTCTCGCGCGC---GACGACTCCGAGGCCATCCCGCGCG--TCGGCTGTTC 257626
Qy 2367 aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
Db 257627 GGGATCTAGGTTTCGACTCGTTCGGCGGTGGAGCTTCGCAACCACTCCGACACAGA 257686
Qy 2387 pGly-----ArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLy 2404
Db 257687 CGGAGCTGGCTCTCCGACGACTCTCGTCTTCGATTACCCGAGCCCAACCAAGCTCGCC 257746
Qy 2404 sValSerGlyArgProSerSerArgLysAlaLysSerProAla----- 2418
Db 257747 AATTCTGCTCT-CCGAGATCGCG---AGTTCAGCGCCGCAACTCAACTCCGCTTCCG 257802
Qy 2419 ----ProGlyLeuAlaSerGlyAspArgProProSerValSer---SerValHisSerGl 2436
Db 257803 CGACCCCGCGGAGCTCGATGAGCCGATCGCCATCGTTCGATCGGCTGTGCTGCTCCC 257862
Qy 2436 uGlyAspCysAsnArgThr-----ProLeuThrAsnArgVa 2449
Db 257863 GCGGAGTGACCTCGCGGAGGACTTCTGGGATCTGATCTCTCGAGCAGGACGCGATC 257922
Qy 2449 lTrpGluAspArgProSerSerAlaGly-----SerThrProPheProTy 2464
Db 257923 GCGGATTCCTCCACCGACCGCGGTGGGACCTGGACACGCTCTACGACCCCGACCCGAC 257982
Qy 2464 rAsnProLeu----- 2468
Db 257983 CACCCCGGCGCTCTACACCCGAAACGGGGATCTCTTACGAGCAGGCGCACTTCGAC 258042
Qy 2468 eMetArgLeuGlnAlaGlyValMetAlaSerProProPro-----ProGlyLeuProAl 2486
Db 258043 GCCGAATTCCTGGGCATCAGCCCCCGGAGACCTCTCGCATGGACCCCGACCAAGACTC 258102
Qy 2486 aGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluGluProLysPro 2504
Db 258103 CTCCTCGAAACCGCTCGGAAACCATCGAACACCGCGCATCAACCCCGCACCC 258157

RESULT 80
LOCUS AE006968/c 16188 bp DNA linear BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 54 of 280 of the
complete genome.
ACCESSION AE006968 AE0000516
VERSION AE006968.1 GI:13880290
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1 (bases 1 to 16188)
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Uterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL Unpublished
```

REFERENCE
AUTHORSTITLE
JOURNALFEATURES
source

```
1..16188
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/notes="clinical strain"
111..1436
/gene="MT0756.1"
111..1436
/gene="MT0756.1"
/notes="similar to GP:2911815; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="preprotein translocase SecY subunit"
/protein_id="AAK44990.1"
/db_xref="GI:13880291"
/translations="MISAFISLSRLTDLRRKILFTGLGILYRVGAALPSGVNPPNV
QQICKASAGEAGQYSLINLFSGALLKLTFAVGMPYITASIIYQLLTADQISFTL
ELRKGAGQSKMTQYTRYLAIALAQLQATSI VALAANGLLQCCSDITADQISFTL
VTVLVMTGAALVMMGELITGERIGNMSLIFVGAIRI PAEGSILDESIGTFV
TAVCAALIIIVGVFVEQQRRI PQYAKRMGRMYGTTSTYLPKVNQAGVPI
FASLIYIPLHITQIRSGVSGVSMWDFVGTYSDPNLTIGIYFGIIEFTYF
VYSITNPDRDEAMKKFGGFI PGIRPRTADYLRVLSRTILPGSIYLGIVALPN
LPQIGAGCTVQNLPGGTAVLIMIGVGLTVKQIESQLMQRNYEFLK"
1433..1978
/gene="MT0757"
1433..1978
/gene="MT0757"
/notes="similar to PID:2344858; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="adenylate kinase"
/protein_id="AAK44991.1"
/db_xref="GI:13880292"
/translations="MRVLLGPPGAGKGTQAVKLAELGIPQISTGELFRNIEGTK
LGVEAKYLDAGDLVPSDLTNELVDRLNPNDAANGFILDGYPKSVQAKALHEMLR
RGTDIDAVLEFRVSEVLLERLKRGRADTDVILNRMKYRDETAPLLEYRDLK
TVDAVGTMDVFAFARALALGK"
1981..2781
/gene="MT0758"
1981..2781
/gene="MT0758"
/notes="similar to GB:D00619 SP:Pi9994 PID:216341
PID:1044986 GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="methionine aminopeptidase"
/protein_id="AAK44992.1"
/db_xref="GI:13880293"
/translations="MRPLARLGRVRVVPQSRAGELDAMAAAGAVAAALRAIRAAAP
GTSLSLDEIAESVIRESGATPSFLGVHGPASTCASINDRVVHGIPTAEVLAPGDL
VSIDCGNALDGHGDAAITPGVGLSADADALSEATRESLQAGIAMVGNRLDVAAH
AIEGTAAELRYSGFVIGVAGGHGIGRQMDPFLPNEGAPRGPGLAAGSVLAI
EPMLTIGTKTVVLDKMTVTADGSRRAHWEHTVAVTDDGPRILTIG"
2854..3387
/gene="MT0759"
2854..3387
/gene="MT0759"
/notes="similar to GP:5834669; identified by sequence
similarity; putative"
```

```
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-70 factor, ECF subfamily"
/protein_id="AAK44993.1"
/db_xref="GI:13880294"
/translation="MARVSGAAAEALMRALYDEHAALVRLRYALRLTGDAQAQADV
QETLRWQHEVIGDITAPARAWLFTVARNMIIIDERSARERNVVGSTDSQGTPEQS
TDEVNALDRLLIADALQSAEHRRAVIRQSYRGSWTAQIATDGLGAEGVKSRLLH
YAVRALRLTQLQELGVR"
3451..4203
/gene="MT0760"
CDS
3451..4203
/gene="MT0760"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44994.1"
/db_xref="GI:13880295"
/translation="MTMPLRGGLGPPDDTGVREVSIGDDHHYAMWDAAYVLGALSADR
REPEHLACGECRGAVTELCGFALLSQLDRDEVAIISAPTVVASGLSPILLPSL
LAATHRRRLRLITWVASSAAALVLAIGLVQVGHSAAPQRAAVSALPMAQVGTQL
LASTVISGEPWCTFINLRCVCLAPPYASHDTLAVVVGSDGSTRLATWLAEPGHTA
TPAGSISTPVDQIAAQVVAADTGQVLLQSL"
4258..4470
/gene="MT0761"
CDS
4258..4470
/gene="MT0761"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44995.1"
/db_xref="GI:13880296"
/translation="MVGRPPPARSLHDAHLQSSITVVLAKEHFSAAATVADGYRSRG
AGFVAAASGGGFTLQKCAATAS"
4518..5015
/gene="MT0762"
CDS
4518..5015
/gene="MT0762"
/notes="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, MarR family"
/protein_id="AAK44996.1"
/db_xref="GI:13880297"
/translation="MASDNRDPIAARANWERSGWDVSLGMVAVTSVWRAHOILLAR
VETALRPYDLSPFSFELLRLIAFSRIGALPITKASDRLOVHTSVTHAIRLEADGLV
RVPHPTDGRITLQITELGRSTVEDATVTLNEQVFANVGMAEESQALVSAVETLRR
NAGDF"
5012..5224
/gene="MT0762.1"
CDS
5012..5224
/gene="MT0762.1"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44997.1"
/db_xref="GI:13880298"
/translation="WRAGRRVSAQCRAEMGAYASARALGAQPATFHDQFVRLTRAR
ARMRPPRAGLLNGGARRYPAGS"
5373..5921
/gene="MT0763"
CDS
5373..5921
/gene="MT0763"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44998.1"
/db_xref="GI:13880299"
```

```
/translation="MDPLMAHORAQDAFAALLANVRADQLGGTTPCSEWTINDLIEHV
VGGEOVQRWAASPIEPARPDPGLVAARHAAVAHEIFAAPGMSATFKPLGCEVPG
QVFTGLTQVLTWHLAAATQSTDLDELAVERLAAARALVGPFRGPGKPFADDE
KPCPRPPADQLAAFLGRVVR"
6166..6972
/gene="MT0764"
CDS
6166..6972
/gene="MT0764"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44999.1"
/db_xref="GI:13880300"
/translation="MVLTRAREVALTQHIGVSAETDRAVVVKLRQAYDSILVCGRRRL
CAIGAEINAVAHQORALGLDTPAGARNFSEFLAKHDIITRLVLAATAESAGARLR
SLASYQAVGFGPKPQPPDPVFPFPYQPKWAACRARGODPKVVTTFHAPMSAR
FRSLPAGDSVLVYCGNDKYGLLHIQAKHGRQWHDIAARWPSAGNWRVLAIDYAGNTLA
YPERVEYNQNDITFAVYRMSLPDGRVYVFTTRVLIISARDGKLIITAFPTT"
7006..7254
/gene="MT0765"
CDS
7006..7254
/gene="MT0765"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45000.1"
/db_xref="GI:13880301"
/translation="MTGPPRSYTGRRDLIAEKLPYFOISAMLPKNTPTSETAEFFW
DNSLWCWSGDRGTGYTRITVTVSIQVADGERAEFGT"
7251..7613
/gene="MT0766"
CDS
7251..7613
/gene="MT0766"

Alignment Scores:
Pred. No.:          Length:      16188
Score:              684.00      Matches:      498
Percent Similarity: 32.80%      Conservative: 162
Best Local Similarity: 24.75%      Mismatches:    802
Query Match:        5.18%      Indels:        559
DB:                  1          Gaps:            85

US-09-522-753-5 (1-2517) x AE006968 (1-16188)
Qy 668 GlnHisLysLeuLysMetGluLysGluArgAsnAlaAraArgLysLysLysLysAlaPro 687
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
16186 CAGCATCATCTCGGTC-GAGCGCGCGAAGCATGCTGTCAACGCCGCGCGCGTGCAG 16128
Qy 688 AlaAlaSerGluAlaAlaAlaPheProValValGluAspGluMetGluAla 707
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
16127 GCCCGCGCGCTACATCAAGCGCTTCAGCCGCTTCTCACCGAGCAGCGCCGCCGCCGCC 16068
Qy 708 SerGlyValSerGlyAsnGluGluMetValGluAlaGluAlaLeuHisAlaSer 727
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
16067 GCCGCGCGCGCTCGG-----GCCGGGTGCACCCCGCATTTCCCGCGCTTACC 16053
Qy 728 GlyAsnGluValProArgGlyGluCysSer-----GlyProAlaThrValAsn 743
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
16052 -----GCCGGGTGCACCCCGCATTTCCCGCGCTTACC 16011
Qy 744 AsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAla----- 759
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
16010 GCCGTTGCCCGCGTTCGCCCGCTTGCAGATCAGCAGCGGCTCCCGCCGCGCCGCTT 15951
Qy 760 -----LysAspThrGlyGlnAsnGly-----ProLysProProAlaThrLe 773
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
15950 GCCCGCGATGTGGAACACCGGACAAACCGGTGGCGCGCGCGCGCGCTCACCTCC 15891
Qy 773 uGlyAlaAspGlyPro-----ProProGlyProProThrProPro-----ArgArgTh 789
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
15890 GAGACCGAAGCTCCGACTTGCAGCGCGCGCGCGCGCGCGCTCCCGCGCGCGCGCT 15831
```


Db 14115 CGCGGTGGCGATACAGCGGCGCGCGTGGCGCGCCACGAATTCGCGGTGTGATCGGGGCCA 14056
QY 1458 --LysTyrAspThrGlyAlaSerThrThrGlySer-LysLysHisAspValArgSerLeu 1476
Db 14055 GCAGCGCGACACGGCGCGCTCGCGCGCGCATACGCGCTCGCGCGGCTCAGGG 13996
QY 1477 IleGlySerProGlyArgThrPheProValHisProLeuAspValMetAlaAspAla 1496
Db 13995 CTTGACGAACTGGCGGTAAACGCGCGCG-----CCTGTG----- 13960
QY 1497 ArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGly----- 1513
Db 13959 -----CGTCCGCCCTGGTAGGCTGGCTTGGGCAC 13927
QY 1514 ---ThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGlu 1532
Db 13926 CGAACACCGCGCAATCGCTGTCGACACCTCATCGGCACCGCGGTCAACAATCCGTCG 13867
QY 1533 LeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGly 1552
Db 13866 TGTGACGGCG-----CGCGCG 13849
QY 1553 HisLeuPro-----ArgGlySerProValThrMetArgGluProThrPro 1567
Db 13848 CATTAGCGGAGCTAATCGCGACCGCATCCCGCGCAATCCGCTGCGCGGCCACCA 13789
QY 1568 ArgLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuThrSer 1587
Db 13788 GCTCGGCGAAACCATCACCATGACATCGCATCGCTCCCA-----CCCTCAGCGCC 13735
QY 1588 -----ThrProArgGluIleAla----- 1593
Db 13734 TGTGCGCGACTACCAGCACACACCGTAGCGGTAGGAATTAATGCAGAAATCAACGA 13675
QY 1594 -----LysSerProHisSerThrValProGluHisHisProHisProIleSerPro 1610
Db 13674 TATCGCATCGGACCGCATCGAGCGCTTCCACACGTTTCGCAG-----GCGCCA 13624
QY 1611 TyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIlePro--- 1629
Db 13623 GCTGCTGAGCGGCTCAGCGGTGCGCGCGGTTCGCCACAAACGGCGCGGTGCGCGGC 13564
QY 1630 -----LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAla 1647
Db 13563 ACAGGGCGCGCGGAACCCGCA-----CCGCGCGATAGCGTCTCAGCAGACCC 13510
QY 1648 AlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr-ProPr 1667
Db 13509 GCG-----ACGGGCTCTGACAAAGCGCTCGAGCGCGGGGATCTTGGCGGAG 13462
QY 1667 oTyrIleuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAs 1687
Db 13461 ATACCTGCGCGCGCG-----GCCCCCATTTTGGCGCGCGCG----- 13426
QY 1687 nAspTyrIleThrSerGlnGlnMetHisAsnThrAlaThrAlaMetAlaGlnArgAl 1707
Db 13425 -----TCCGAGCGCATCTGCATCAGCGGAGTGGTTGACCGCGGAGCATG 13376
QY 1707 aAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaG 1727
Db 13375 CCGTGCGCGCGGCTACGCAACCGTTGAGCCCGGTTTTGGCCCGCAGCAACCGCGCG 13316
QY 1727 Y-Pro-----ArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValP 1745
Db 13315 CGCGCGCGGTTCGCGGCTGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTG---C 13259
QY 1745 roProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlap 1765
Db 13258 CGCGGTGCCGATCAGCGCGCGCGGTGCGCG-----CCGACGCCAC 13220
QY 1765 roGlnProPheSerSerArgHisSerSerProLeuSerProGlyGlyProThrHis 1784
Db 13219 CGGCGCGCG-----CCGAAGCCAGCCAGCGGCGATTCGCGCGCGCGCGCGCGCGCGCG 13166

QY 1785 -----LeuThrLysProThr-----ThrThrSerSerSerGlu 1795
Db 13165 CGCGGTGCCGAACCGCGCGCAACCGCGCGCGCGCGCGCTGACACAGACCGCGCGGT 13106
QY 1796 ArgGluArgAspArgAspArgGluArgAsp-----ArgAspArgGluArgGluLys 1812
Db 13105 TGCCACCGCGCGCGCGCTCCACCGAGAAACCAATCCCGCGCGCGCGCGCGCGC 13046
QY 1813 SerIleLeuThrSerThrThrThrValGlu-----HisAlaProIleThrArgProGly 1830
Db 13045 CGCTCCCAACAGCAGCGCTGCGCTACCGCGCGCGCGCGCGCTCCACCGGTGTCCAGT 12986
QY 1831 ThrGluGln-SerSerGlySerSerGlySerSer-----GlyGlyGlyGlyGlySerSe 1848
Db 12985 TTTCGCGGTACCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCATTCGCGCATTC 12926
QY 1848 rSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrG 1868
Db 12925 CGCGCGCGCGCAACCGCGCGCTGCTAGTCTGAGCCCGCGGTTCGCGCGGTGAC 12866
QY 1868 nAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleth 1888
Db 12865 CTCCGCGCGCACCGCATCAGCGCGC-----CAGCAGTCCCGCGGTCCCGCGCGC 12817
QY 1888 rAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSer-ProVala 1908
Db 12816 -----CCACCGCGCGCGCGCTC-----GCCACCAACAGCGCGGTTC 12779
QY 1908 rGProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyV 1928
Db 12778 CACCG 12741
QY 1928 aTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgP 1948
Db 12740 -----CCGACCGCGCGCTCCCGCGCGGTGCCCGCGAGT-----ACGCTGCTGGCTC 12695
QY 1948 roGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerG 1968
Db 12694 CGCGCATCCG 12641
QY 1968 lYLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValProPro 1988
Db 12640 CC---CCACCG 12584
QY 1988 aSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlas 2008
Db 12583 TACCC---CACAGCGCG---GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12530
QY 2008 erProAspProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnS 2028
Db 12529 CG 12489
QY 2028 erLysProPheSerIleGlnGluLeuLeuArgSerLeuGlyTyrHisGlySerSert 2048
Db 12489 ----- 12489
QY 2048 yrSerProGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspL 2068
Db 12488 -----CCGGGCGCGCGGTGCGCG-----CCGCTGCCAGCGCGCG 12455
QY 2068 ysGlyLeuProLysHisLeu-GluGluLeuAspLysSerHisLeuGlyGlyGluLeuArg 2087
Db 12454 CGGGGTGCGCG-----TTGCCGATCAGCGGTCCCGCGGTGCTGTGGGTGCGCGCGT 12401
QY 2088 ProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArg 2107
Db 12400 TGATCGCGCGCGCGCGCTGTT-----GTTTACGGGTGTGACGCGGAGAG-GCGCTGGCGCG 12345
QY 2108 ProLeuProGluSerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLys 2127
Db 12344 GCGTTGAAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12294

Qy 2128 GlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyr 2147
Db 12293 -----GTATACCCCGGTCGGTCGGT----- 12270
Qy 2148 ThrArgHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhePro 2167
Db 12269 -----CCGCGATGCGCGGTCGGTCGGTCGGTCGGTCAGACCCCGTCGCGCGG 12219
Qy 2168 GlyAlaSerCysProValLeuAspLeuArgProPro----- 2180
Db 12218 GCGCGCGCGGTACCGCGGTACCGCGCTTCCACCGCGCGCGCGCGCGCGGTTA 12159
Qy 2181 -----SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPro 2198
Db 12158 CCGATACGACAAACCGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 12099
Qy 2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVal-----Leu 2215
Db 12098 AACCG 12040
Qy 2216 GlyGlyGlyGluAspGlyIle---GluProValSerProProGluGlyMetThrGluPro 2234
Db 12039 GCGCGCTCCCGTCGGTATCGCGACCCCGCGTCCCGCGGATCCCGGTCGGGTCGCGAA 11980
Qy 2235 GlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluInThrGluPro 2254
Db 11979 CAGCACACCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11944
Qy 2255 Ser-----ArgMetGlySerLysSerProGlyAsnThrSerGlnPro 2268
Db 11943 ACCGTACACCTTAGCGCGGTCCGCGGATCCACCGGTGCGCGCGCGCGCGCGCGCG 11884
Qy 2269 ProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGln 2288
Db 11883 CCGCGGTGCG 11851
Qy 2289 GluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGln 2308
Db 11850 ---GTCGTTGAGAAAGCACGTCGCGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCG 11794
Qy 2309 ProGly-----ThrGluPheAsnMetProAlaIleThrGlyThrGlyLeu--- 2324
Db 11793 CCGCGCGCAGCAGCCACCGCGGTCCGCGGATCCCGCGCGCGCGCGCGCGCGCGCG 11734
Qy 2325 -----MetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGly 2341
Db 11733 CCG 11680
Qy 2342 LeuGluAlaIleLeuArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluSerPro 2361
Db 11680 ----- 11680
Qy 2362 ProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro-----Ala 2379
Db 11679 CCG 11620
Qy 2380 AlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeu----- 2394
Db 11619 CAACAACAGCACCTGACCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11560
Qy 2395 -----ThrSerProGlyGly-----GlyGlyLysAlaLysValSer 2406
Db 11559 CCG 11500
Qy 2407 -GlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspAr 2426
Db 11499 GGGCAGCG 11443
Qy 2426 gProProSerValSerValHisSerGluGlyAspCysAsnArgThrProLeuTh 2446
Db 11442 CCG 11398
Qy 2446 rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhe---ProTyrAs 2465

Db 11397 GCCGTTCGATCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11338
Qy 2465 nProLeuIleMetArgLeuGlnAlaGlyValIleAlaSerProProProGlyLeuPr 2485
Db 11337 CAGTCTCTGTCGACGCTCTGCGCGCGGTTCGCGACGACCGCGCGCGCGCGCG 11279
Qy 2485 oAlaGlySerGlyPro 2490
Db 11278 CGCTGGCGTAGGCGCC 11263

RESULT 81
AC087386/c
LOCUS Homo sapiens chromosome 15, clone RP11-173D3, complete sequence.
AC087386
AC087386.7 GI:24111062
VERSION HTG.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

REFERENCE
1 (bases 1 to 155343)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173D3
Unpublished
2 (bases 1 to 155343)

[illegible]


```
QY 2407 yArgProSerSerArglysaAlaLysserProAlaProGlyLeuAlaSerGlyAspArg-P 2427
Db 77523 -----GGGTCTTACCCGGTCTCGACC 77545
QY 2427 roProSerValSerValHisSerGluGly-----AspCysAsnArgArgT 2443
Db 77546 CTAGCTTATTGGGTGAATAAGACTCAGATGCCCGCTGCTCGGAATGTCTTACATTGA 77605
QY 2443 hrProLeuThrAnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheP 2463
Db 77606 CGGAATAGTGGCTGAAGTTGGTGTGGCTGGCGCTTAACCGGGCAGCATCCCT---- 77661
QY 2463 roTyraAnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProG 2483
Db 77662 -----CCCTGCTCTCCACCCTCCCG 77683
QY 2483 lyLeuProAlaGly-----SerGlyProLeuAlaGlyPro----- 2494
Db 77684 GGCTGGACCCGGCGGTGGCTGGCGGGGTAGGATTGGTCCGGTAGGTGTG 77743
QY 2495 --HisHisAlaTrpAspGluProLysProLeuLeuCysSerGlnTyxGlu 2511
Db 77744 CGCGGCATCTCTGGCGCAGT-----TGTTCCCGTTACGAA 77778

RESULT 83
SC0939119/c 299050 bp DNA linear BCT 11-FEB-2003
LOCUS SC0939119
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 16/29.
ACCESSION AL939119 AL353816 AL356612 AL357432 AL357524 AL358692
AL358818 AL359988 AL391406 AL391541 AL392149 AL392150 AL392178
AL645882
VERSION AL939119.1 GI:24427855
KEYWORDS
SOURCE
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1 Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabinowitz,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrall,B.G.,
Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21996410
PUBMED 12000953
REFERENCE 2 (bases 1 to 299050)
AUTHORS Bentley,S.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk
COMMENT On or before Oct 30, 2002 this sequence version replaced
gi:7242748, gi:7636006, gi:8248766, gi:8249971, gi:8388712,
gi:8439477, gi:8894718, gi:9857143, gi:9886716, gi:10129746,
gi:11595528
FEATURES
source
1. .299050
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
200..1727
/gene="rrnA"
200..1727
/gene="rrnA"
/note="rrnA, 16S ribosomal RNA. Coordinates based on

comparison with rrnB from Streptomyces coelicolor
(EMBL:AL109848)."
2004..5126
/gene="rrnA"
2004..5126
/gene="rrnA"
/note="rrnA, 23S ribosomal RNA. Coordinates based on
comparison with rrnB from Streptomyces coelicolor
(EMBL:AL109848)."
5228..5348
/gene="rrnA"
5228..5348
/gene="rrnA"
/note="rrnA, 5S ribosomal RNA. Coordinates based on
comparison with rrnB from Streptomyces coelicolor
(EMBL:AL109848)."
complement(5412..5680)
/gene="SCO4123"
/note="synonym: SCD72A.09c"
complement(5412..5680)
/gene="SCO4123"
/note="SCD72A.09c, possible two-component system response
regulator, len: 222aa; similar to many eg. TR:O86632
(EMBL:AL031155) possible two-component regulator from
Streptomyces coelicolor (222 aa) fasta scores; opt: 973,
z-score: 1128.5, E(): 0, 69.1% identity in 220 aa overlap.
Contains Pfam match to entry PF00072 response reg.
Response regulator receiver domain, Pfam match to entry
PF00196 GerE, Bacterial regulatory proteins, luxR family
and Prosite match to PS00622 Bacterial regulatory
proteins, luxR family signature."
/codon_start=1
/transl_table=11
/product="putative two-component system response
regulator"
/protein_id="CAB92376.1"
/db_xref="GI:8052435"
/db_xref="GOA:Q9KYE7"
/db_xref="SPTREMBL:Q9KYE7"
translation="MTVLIADDDPLQRFGRMLLESQDDMTVVGEAANGAEARMAA
ELHPDVLMVDRMPGLDGEATRRIVAGGSRVLLITFDLDEYAGLGAAGSGL
VKDLPPELLSGIRAVATGDVAVPSLTRLRLDAYVHHLPAAGTRDHTPGSPPLRSP
LTEREILRVNGQGQWNTIEARLHAEISTVKTHVGRILAKTGSRDVQAVLAYET
KLVRG"
complement(5415..5609)
/gene="SCO4123"
/note="Pfam match to entry PF00196 GerE, Bacterial
regulatory proteins, luxR family, score 61.00, E-value
2.6e-14"
complement(5475..5558)
/gene="SCO4123"
/note="PS00622 Bacterial regulatory proteins, luxR family
signature"
complement(5736..5607)
/gene="SCO4123"
/note="Pfam match to entry PF00072 response reg, Response
regulator receiver domain, score 126.30, E-value 5.7e-34"
complement(6077..7447)
/gene="SCO4124"
/note="synonym: SCD72A.10c"
complement(6077..7447)
/gene="SCO4124"
/note="SCD72A.10c, possible two-component sensor kinase,
len: 456aa; similar to many eg. TR:O86631 (EMBL:AL031155)
putative two-component sensor kinase from Streptomyces
coelicolor (429 aa) fasta scores; opt: 929, z-score:
995.1, E(): 0, 42.7% identity in 422 aa overlap. Contains
Pfam match to entry PF00512 signal, Histidine kinase and
possible membrane-spanning hydrophobic regions. Contains
possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative two-component sensor kinase"
```


QY 230 IleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGly 249
Db 123039 -----CCGATGACGTCGGCGACGAGGTGCTGGTCGCTTCGACCGGGGGCGCTGGA 122986
QY 250 ProGlnValGluLeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsn 269
Db 122985 CCACCGCTTCGT-----GATCGGGCGCTCTACACCGCGC 122950
QY 270 IleLysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis 289
Db 122949 GGACGTCCCGACCA-----GAGCGAGTCGCGCT 122920
QY 290 AlaArgLysGlnTrpLysGln----- 296
Db 122919 GCACGAGCGCTGAAGAAGAAGCGCGTCGCGCACACCTGTCGACCGCGCAGGCAACCG 122860
QY 297 -----LysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuLysLys 312
Db 122859 GGTGACCTGCTCAGCGACGCGACCGCGCGCGCAAGCGGCGTTCGGATCCCGACGG 122800
QY 313 ValGluArgIleGluAsnAsnPro-----ArgArgArgAlaLysGluSer 327
Db 122799 CAACGACAGCTGACCATCACTCGACCGCACCAAGACCGAGATCACCTCGACAGCAA 122740
QY 328 LysValArgLysGlnTyr----- 333
Db 122739 GGGCTCGCTAGCATCACCGCAGCGCTCGGTGTCGTGCGCGCGCACCGACCTGTC 122680
QY 334 ---GluLysGln---PheProGluIleArgLysGlnArgGluLeuGln-----Arg 349
Db 122679 GCTGAGCGCCAGACGCTCCCTGACCATCAGAGCGCGGTCCCTCAGCATCCAGGCGG 122620
QY 350 MetGlnSerValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGlu 369
Db 122619 CAGCAGCATCAACATCGCAGCGCGGCTGTCGAGTCAACGCGGAGCGCGCTGAA 122560
QY 370 HisGluValSerGluIleAlaAspGlyLeuSerGluGlnGluAsnLeuLysGlnMet 389
Db 122559 CCTGGGCGCGCGCGCGCGC-----GCTGAGCGCGGGCGCGCAGCATCTC 122506
QY 390 ArgGlnLeuAlaValIleProMetLeuTyrAspAlaAspGlnGlnArgIleLysPhe 409
Db 122505 CGCACCGTCAACGTGCGAGATCACTCGTCATGACGGGTGTCATCGCGC----- 122455
QY 410 IleAsnMetAsnGlyLeuMetAlaAspPro----- 419
Db 122454 --GACCTTCAGGT-----CAAGACCCCATCTTCAAGTTCGCGACGATCCCGCTGCC 122404
QY 420 -----MetLysValTyrLysAspArgGlnValMetAsnMet----- 431
Db 122403 GGGCCTGTGACGAGCGTTTCGAGGAGGAGGAGTTGC--TCCTGATGGCGGACAGTTTC 122345
QY 432 -----TipSer----- 433
Db 122344 GTCGCTCGGCTGCTGTTCCCTCGCATCGCGCGACCGGGGATGCGCTCGTC 122285
QY 434 -----GluGlnGlu---LysGluThrPheArg-----Glu 442
Db 122284 AGCGCGAGCAGCAGGTGAGGAGGCGCATGCGGCTCATCTCGCCACCGCGCGGGGAG 122225
QY 443 LysPheMetGlnHisProLysAsnPheGly-----LeuIleAlaSerPhe 457
Db 122224 CGGCGATCGCG-----CCGAGTTTCGCTCGGCATCCAGCATCTGCTTCGCGCC 122171
QY 458 LeuGluArgLysThrValAla-----GluCysValLeuTyrTyrLeuThrLysLys 475
Db 122170 GTCAAGCAGCAGCGCGCGCATCCAGCAGGAGGTATGTACCTCGACCGCTG- 122112
QY 476 AsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgGlySerGln--- 494
Db 122111 GGAACACCATCGAGGTCCACGAGTCACTCACCGCGGAGGACAGAACGTCT 122052

QY 495 -----GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506
Db 122051 CTTTCATCAGCTCGCTTACTCGATCCGCGGACCAACCAACCGCGAGCTCGTCTTCCC 121992
QY 507 GlnGlnGlnGlnProMetProArgSerSerGlnGlnGluLysAspGluLysGluGlu 526
Db 121991 GTTTCATCGCTCATCTCCCTCCACGAGCGAGCTCCCGCAGCGCGCGCGCTCCC 121932
QY 527 LysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAsp---Leu 545
Db 121931 GGGCTCTCCGAAAGCGACCGCTGATGCTCCCTTCCCTTCCACCTCGACGCGCGC 121872
QY 546 LeuLysGluLysThrAspAspThr--SerGlyLysAspAsnAspGluLysGluAlaValAl 565
Db 121871 TTCAGAGAGTTCGTCGACGATCCCAAGCGCTATATCCAGCAGCGCGCGCGAGTGGAC 121812
QY 565 aserLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIle-----Th 583
Db 121811 GACCACACGCTCTCGACCGCGCTGTCGCTGTCGAGACGCTGCGCCACATGCGCCGAC 121752
QY 583 rArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGl 603
Db 121751 CAGATCGTCTACCGCTCAACCGGTCCCGGACAGAACACCAC----- 121709
QY 603 uLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAl 623
Db 121708 -----TCGCTTCTCGACCTCGCTCGCATCATCGCTGTC 121674
QY 623 alysLys--GlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlyS 643
Db 121673 CGCGCTCGCGC---CGCGCGCAGCGAGCTTCTGCTGTC---CGCGCGCAGGAG 121620
QY 643 erLys-----ThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArg--- 659
Db 121619 GACCGATCTCTCGCTCGCGTCGCGCAGGTCGCCACCTCGCGACCGACGACGACGAG 121560
QY 659 ----- 659
Db 121559 GCGCTCGTGTTCGCCACCGCAGCAGGACCTCGCGATCGTCCGTGACAGTGGCGCGCTG 121500
QY 660 -----GlnAsnLeuAspGluIleLeuGlnGlnHisLysLysLeuLysMetGluLysG 676
Db 121499 GTACCGAGTTCAGCGCGCGCGCGCTCAGCGACCGCACCGACCTCGCGAGAGCAAG 121440
QY 676 luArgAsnAlaArgArgLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaP 696
Db 121439 GACGTGCTGCTTCGCG--AGCGCGCAATCCCGCGCATGCTCATCGCGCTCAG 121381
QY 696 heProProValValGlu-----AspGluGluMetGluAlaSerG 709
Db 121380 CGCGCGCTG--CGGACTCGCGCTCGCGCTCGAACTCGACCGCGCTCGCGCGCTG 121322
QY 709 lyValSerGlyAsnGluGluGluMetValGluGluAlaGluAlaLeuHisAla---SerG 728
Db 121321 GTGTGACCGCGCGCGCGCGCTGCTGCTGCGGAGGCTTGAGCGAGGACGGTGGCAGT 121262
QY 728 lyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAsp 748
Db 121261 CCTCGAGGTTCGCGC-----GACCGCACCGGTGCTCTCAACCGCGCGCGCGAC- 121212
QY 748 hrGluSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProL 768
Db 121211 -----GTCGT--GCTCCACATACCGCGCGGACGTGTCTTCCCGCACCGCGCGCAC 121161
QY 768 ysProAlaThrLeuGlyAlaAspGlyPro-----ProProGlyProProThrP 785
Db 121160 GAAG--CGGCTGATCCGCTCGCGGTCCCGGTCACCGAACCGCTGAGCGCGCGCGCTTCTACAC 121102
QY 785 roProArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyA 805
Db 121101 CACCTCGCGACCATCC---GTTCCCGCGAGGCTCACCATCGCGCGCGCGCGCGCTC 121045
QY 805 laProThrProProAlaProProSerProSerAlaProProProProValProLysG 825

QY	1465	rThrGlySerLysLysHisaspValArgSerLeuIleGlySerProGlyArgThrPhePr	1485
Db	119292	-----AGCGCGGTGGCGCGCCGACCG-----CGCGCGCGCG	119261
QY	1485	oProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluAlaGlyCysTyrG1	1505
Db	119260	GCCGCCACGTCCAGCAGCGCCCGGCTTCGACGCTTCGCGGGCGGGGTGCCCGCC	119201
QY	1505	uGlySerLeuLysSerArgPro-GlyThrAlaSerSerSerGlyGlySerIleAlaArgG	1525
Db	119200	GCCGCCACCGCGCGCGCCGACCGGACCGCGCGACCGGACCGCGCGCGCG-----CATCGGG	119144
QY	1525	lyAlaProValIleValProGluLeuGly-----	1534
Db	119143	GAGAACCCGACGCTGCCCGCCAGCAGC-GGCGTAGCGCGCGCGACCGCGGACAGCCCGCGCG	119085
QY	1535	--LysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisL	1554
Db	119084	AGACCGCGCGGCCACACCGCGGACACACCGCGGACCGGGAACACCGCGGCGCGCGCC	119025
QY	1554	euProArgGlySerProValThr-----MetArgGluProThrProArgLeuGlnG	1571
Db	119024	CGCGCGGGAGCGCGCGGACCGGACCGCGCGCGCTCCGGAGCGCGCGCGCG-----	118975
QY	1571	luGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgG	1591
Db	118975	-----	118975
QY	1591	luIleAlaLysSerProHisSerThrValProGlu-----HisHisProH	1606
Db	118974	-----GCCGCGCGCGCGCGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCG	118920
QY	1606	isProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgS	1626
Db	118919	CGCTCCGCGCGCGCG-----	118906
QY	1626	erHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaA	1646
Db	118905	-----CCCGAACCGGCTCCAGACACATGTCGACCGGG-----CCGCTCGCTGTCT	118857
QY	1646	laAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrP	1666
Db	118856	CGTCCCGCTGCCACCGCGGAGC-----CCGCGCGCGCGCGCGCGCGCGCGCG	118815
QY	1666	roProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleI	1686
Db	118814	CCCCG-----CCGTGCGCGCGCGCTCCTCGCGCGCGCGCGCGCGCGCGCGCG	118774
QY	1686	leAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnA	1706
Db	118773	-----ACGCGCGCAGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	118728
QY	1706	rgAlaAspMetLeuArgGlyLeuSerProArgGlySerSerLeuAlaLeuAsnTyrAlaA	1726
Db	118727	CGGCGCGCGCGCGCGGTGTCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	118668
QY	1726	laGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProp	1746
Db	118667	CTGGCGCA-----TGCCCGCTGTCTCGAGCGCGCGCGCGCGCGCGCGCGCGCG	118617
QY	1746	roThrProGly-----ThrP	1751
Db	118616	CTGTGGCGCGCGGTCTTCGCGGGAGCGCGTCCAGAGACCGCGTGGCGCGCGCGCGTCC	118557
QY	1751	roAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerA	1771
Db	118556	CGGCTGCGCGCGGTTCGACCGCATCGGACCTGGATCATCCGTGCGCGCGTGGCGCAC	118497
QY	1771	rgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrT	1791
Db	118496	CCTGGCGCGTCTCTCG-----GCTTCATGTACATCCCGACCGCGGTCCAGAACACCGCGG	118440
QY	1791	hrSerSerSerGluArgGluArgAspArgAspArgGluArgAspArgGluArgG	1811
Db	118439	CTTCCCAAGCGCGCGCGCGCTCCCGGACGGCATCAAGGCTCGC-----	118393
QY	1811	luLysSerIleLeuThrSerThrThrValGluHisAlaProIleThrPargProGlyT	1831
Db	118392	-----GCTCCT-----TCCCGCGGCA	118377
QY	1831	hrGluGlnSerSerGlySerSerGlyGlyGlyGlyGlyGlySerSerSerArg-	1850
Db	118376	CAATCCGAGCTCTCATCAGACAGCTCAGCAACACCTGTGTGGGGCGCGGATCACGGA	118317
QY	1851	--ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerPro-----ArgThrG	1868
Db	118316	GTCCGGCAGCGCGCGTGGTGGAGCGAGCTTACCGAGCGCGACCGGCTCTCGACGT	118257
QY	1868	InAspAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleT	1888
Db	118256	GATCATCACCGCGCGGTCTCCACCGCGCGCGGACAGCTGCGGAGTCCGCGCTCCCGCA	118197
QY	1888	hrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerSerProValA	1908
Db	118196	CGCG-----TCAGCGGACGATCACCATGAAGCGGTCGGTCTCA	118155
QY	1908	rgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyV	1928
Db	118154	GACCGCGACCTCA-----CCTGGACCGCGCGCGCGCGCGCGCGCGCGCGCG	118101
QY	1928	alTyrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgP	1948
Db	118100	CG-----TCGCGGAGGTCCACCAAGTCCGCTTCC	118071
QY	1948	roGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaAlaArgSerG	1968
Db	118070	CATCGAGTCCGCCACG-----CGCGCTCCCGCAAGAGCAGGT	118032
QY	1968	lyLeuGluProLaserSerProSerLysGlySerGluProArgProLeuValProProv	1988
Db	118031	GGCATCTCCGAGATCGAGTCTCTTCG-----GCCGTCCAGCGCGCAACCGCTCTCTG	117978
QY	1988	alSerGlyHisAlaThrIleAla-----ArgThrProAlaLysAsnLeuA	2003
Db	117977	ACCG	117918
QY	2003	la-----ProHisHisAlaSerProAspProProAlaProProAlaSerA	2018
Db	117917	CCCCACCGACCGACCGCACCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	117861
QY	2018	laSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluL	2038
Db	117860	CCATGCCCGCGACGACGAGC-GATATCCACCGCGCGTCT	117824
QY	2038	euArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerP	2058
Db	117823	-----GCTGTCCGCGCGCACCTCTGCGCTCTTCTGAGCGCTCGCGCGCGCGCG	117769
QY	2058	roValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuA	2078
Db	117768	CGGACCG	117736
QY	2078	spLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGlyG	2098
Db	117735	ACCACCGACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	117677
QY	2098	lyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerP	2118
Db	117676	-----CTGCCC-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	117637
QY	2118	roLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgValThrLeuAlaGlnH	2138
Db	117636	CC-----CGCGCGCGCTCCC-----CGACGGCACCGCGCGCAATCGCGCACCTCGCG	117589
QY	2138	is-----IleSerGluValIleThrGlnAspTyr-----ThrArgHisHisP	2152


```
Db 117588 ACGCCGTTGGACTGGAAGAAACCCCTCCGGCGAGACCATGGACCTCGCACTCGTCCGGGCC 117529
Qy 2152 roGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysp 2172
Db 117528 CGCACGAGCGCACCCCGGACGACCGGATCGGCTC-----C 117493
Qy 2172 roValLeuAapLeuArgPro-----ProSerAapLeuTyrL 2185
Db 117492 CTGCTCTTCAACTTCGGCGGCGCGCGGCTCCGGCGTCCGACCTTCGG- GCCTGGC 117434
Qy 2185 euProProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGly----- 2202
Db 117433 CGCCGACTACGAGAACTCCGACACCGCTCGACCTCGCTACGCTCGACCGCGCGCGGT 117374
Qy 2203 -----GlyLysArgSerPro---GluProAsnLysThrSerValL 2215
Db 117373 CGGCAACAGCGCGGCTCGCTCCGCGAGCGCGCTCCCGACGAGGACGAGA---- 117318
Qy 2215 euGlyGlyGluAapGlyIleGluProValSerProProGluGlyMetThrGluProG 2235
Db 117317 -----TCGACGGCATCCCCGACGACGCGGGGCGGACGAGACCGA 117281
Qy 2235 lyHisSerArgSerAlaValTyrProLeuLeuTyrArgAapGlyGluGlnThrGluProS 2255
Db 117280 CGCGCTCGCTCGCTCAACC-----GGGAGACCGCGCGCGCTG 117242
Qy 2255 exArgMetGlySerLysSerProGlyAanThrSerGlnProPro----- 2269
Db 117241 CGAGAGAACTCCGGCGAGCGCTCCCTCGCTCGGACACCCAGCGCGCGGGACAT 117182
Qy 2270 -----AlaPhePheSerLysLeuThrGluSer-----AsnSerAlaMetValLysS 2285
Db 117181 GGACCTGATCGCGCAGTCTCTCGCGACGAGAACTGCATCTACCTCGCATCTCTCTACGG 117122
Qy 2285 erLysLysGlnGluIleAanLysLysLeuAanThrHisAanArgAanGluProGluTyrA 2305
Db 117121 CACCGAACTCGCGCGCTCTACGGCGCATTTGTTCCCGGAGCAGCTCGCGCGCGCTACT 117062
Qy 2305 snlle-----SerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyT 2322
Db 117061 CGACGGTGTGTCGACCCGACCCAGGACAGA----- 117030
Qy 2322 hrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyL 2342
Db 117029 -----TGAGGAGAGCCCTCGCCAGCGCGCGGCTTCCAGCTCGCCT---- 116988
Qy 2342 euGluAlaIlelleArgLysAlaLeuMetGlyLysTyrAspGlnTyrGluGluSerProp 2362
Db 116987 --TCGAGCACTTCGCGCCTGTTGCGCGCGGAGGCTGCACACTGGCGGAGGACGCCGA 116930
Qy 2362 roLeuSerAlaAanAlaPheAanProLeuAanAlaSerAlaSerLeuProAlaAaMetP 2382
Db 116929 GGCATCGTGGACCTCGCTCGCGCTCGGAGCGCGCTCGGAGCGCCCTCGGACGACACCC 116870
Qy 2382 rotLeThrAlaAlaAapGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyL 2402
Db 116869 CCGGACGACGCGGAATCG----- 116850
Qy 2402 ysaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuA 2422
Db 116849 -----ACGCGACGCGCTCGTCCGCGCGCTCACCAGCGCGCTCTACCGGCGAGGA 116801
Qy 2422 laserGlyAapArg-----ProPro-----SerValSerS 2432
Db 116800 CTATGCGCGCGCTCGGCTCGGCTCTGGAAGCGCGCGCGACGACGACGCGCGCTCT 116741
Qy 2432 erValHisSerGluGlyAspCysAanArgArgThrProLeuThrAsnArgValTrpGluA 2452
Db 116740 GCAGGACCTCCGGAAGCCCTCGGCGAGCAGCGGACCGGCGGCGGCGGACGAGCG--ACGG 116684
Qy 2452 spArgProSerSerAlaGlySerThrProPheProTyrAanProLeu-1leMetArgLeu 2471
```

VOTVI PARTAWLRDHGVRAWEPAAREROTDRKHSCKDKGKACGCEGIAVLLAWFEBAE
 KLLRELGDVFTGIAQEARASAGVTLVWSQKASGYQLSTDTRASLPAAMCFGKGGDA
 SPALPEVLDADPAAMGNKRGYVILASAGVDELDHATPVKRTVWISGSGYERVAE
 YVVTQETTRAAIDAVTADAATKFAKGFTRDRARAFAPAGAPAGIRLITABODVDQE
 KELVDADADVDPTAAILRPETPLSPPPASKPGPEARELLDELVRLMAGMGFGTVAV
 HDLGPVLEOLGRDRSVSKEMKRLAEBGLVPVLAVG"

gene 3988. .4500

/gene="spdB1SL"

CDS 3988. .4500

/gene="spdB1SL"

/codon_start=1

/transl_table=11

/product="putative spread protein"

/protein_id="BAC21267.1"

/db_xref="GI:23978904"

/translation="MTDRQYPTGRQITDADLAAALLCPTRPAPPAPARPQPG
 PAVDVAALLEFARQGALEAAVTQTAPAAETSASILPRWAVGTAVASVIGIAPCLLG
 WADLLATGAAGAAVAAGESPPADPQCILIGAVLDRQPSIARAPKGGGIEVTQITQTI
 TQSVKGGGR"

gene 4497. .5723

/gene="spdB2SL"

CDS 4497. .5723

/gene="spdB2SL"

/codon_start=1

/transl_table=11

/product="putative spread protein"

/protein_id="BAC21268.1"

/db_xref="GI:23978905"

/translation="MKLDTHAVRSTGRSRLRLVLIWLVFGAVLFSVLTPTPLVVRTP
 EGMWTAWILPFFVVDAAVTVVRVDEITARLDGRAGWPLALRWLTGCMTALNIGDS
 MLKGDWGVGHVAVPTLLIFTAAEALGWERAJTRAVARIESSRAERAGARALSART
 VSAQAKARDARAREQORLADERAREQORHAREQOQALEREHLANDAEQAR
 AEREHAQVERERREDAORREEQORADAKAREQOQALEREHLANDAEQAR
 AEARRAPAVSAVSTPRPTVSTVTPAHDITKTPTAETLKMTEADARRAAVEAREGRSQRQAALTGWVGW
 PRPTVSTVTPAHDITKTPTAETLKMTEADARRAAVEAREGRSQRQAALTGWVGW
 VAKRFODLEOTGVSA"

gene 5720. .6166

/gene="spdB3SL"

CDS 5720. .6166

/gene="spdB3SL"

/codon_start=1

/transl_table=11

/product="putative spread protein"

/protein_id="BAC21269.1"

/db_xref="GI:23978906"

/translation="MSTDDLNAHAEVRAEITRTDTKAGLLIAFVGAVLAGAGSIAHD
 VPMTLPAYIAGGIGCAALLAAAVLLHVPRGLGRGHGWLWATLTPQQLAAEPREP
 GRDVVGLSRLLAAAKFTGLRQAILDLITAGAAFAVAAVLAALAGAA"

gene 8113. .8865

/gene="orf753"

CDS 8113. .8865

/gene="orf753"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAC21270.1"

/db_xref="GI:23978907"

/translation="MKWEDRATTDREKIRAGAQGGPPNVGLATPGSLGVVDLMPKE
 GASLGAEEVAFALCERAGQPLPFRAGTASGGEHLFTAPVGIRLGNAGLLGQLV
 DTRWAGGVVAPGSIPTCGVYTVLDDAPVLPPLPSWLLLEALTAROKSRRPRGGVPAPFA
 VSGSRAARALEREQGNVQAAPTQANNLTNRSFLVGRFVANGDIDRQVVEBAFQGA
 GERGLTAACRATIRSLDSSLRKAPREVA"

gene 9198. .10739

/gene="orf1542"

CDS 9198. .10739

/gene="orf1542"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAC21271.1"

/db_xref="GI:23978908"

/translation="MTAHPDELWAGFDEMSAEDVAGPIWDEPVLNPRGQLPVFPVDA
 LPDLWAAMTAGVAEETQTPVDLAGCLALAVVGTAAGRLTVTVRGQWSEBPVNLVTAVA

LPNGNRKSAVFGMLTNPLLAERKALIELTAPORAEAAASARIKAAAEKAEKLAANAE
 ADKLADLTAAQVSLSEAEERAVIPAEPOLVADVTAESLTTLLAQODGRISILSPGGE
 IPEILAGYSGAPNMGIFLKHAGDMARVNRQARDAQYIENPAITWGLAVQPDVLDAL
 GQIKADGGLLARFLYSRPELSLVGRNLTPELLSPDIADTYAQLGRALITLAGWTE
 TAEFLTTPADAVLILAYQRVTSRRLKDGFLAPIINWASKRGAVARIAGLLHLAHGP
 EDAMTRPTAAADTAAATLGDYFTAHLDVFNAMKADPAOEAAATLLTLTETRTAAF
 TKRDLFRAMPSEFPTMSDLDPALNLLHEHGVWRQPPPTTRGGRRPSPRYETHPR
 ITPAA"

gene 11232. .12608

/gene="int"

CDS 11232. .12608

/gene="int"

/codon_start=1

/transl_table=11

/product="putative integrase"

/protein_id="BAC21272.1"

/db_xref="GI:23978909"

/translation="MGRATDQRRRLRHAPARAARSSGAERAPFYSERIHSTMAGH
 VODRYKTELGSQDGKPRRVKTRFGTGRYARYVGPDTGTEKSKSPDRQKRLAEQWL
 NETAADMARGHYIDPRAARITPROFAEKWVATHTEVNSKEAERLRHLVYPIGTR
 PLASFOGHIRTWLGELESKYPAASHRRIVFGTVSAFGALDDGILAKNPCKARSIO
 IPKPIPTRVPTMTQVFGVRAALPERLRATVDVAGCGLRQGEVGLSEDELDPRGG
 WLSVNHQLXRIIRGKFVFLPKGAKVRDVLPTDVALQEHQRQKYPSVAVTLPMWTPD
 GLPVTKSLIFSGSTRDHVRVSHFNHMKPALAAVGIIPELVGKRSPSEHGMHALR
 HFVASVLLDAGENIRALSQYLGHADPGFTLRTVTVTHLMPSEGRTRKAIESLYQQLRDH
 DHGPETAQ"

gene 12990. .13451

/gene="korSL"

CDS 12990. .13451

/gene="korSL"

/codon_start=1

/transl_table=11

/product="putative transcriptional regulator protein"

/protein_id="BAC21273.1"

/db_xref="GI:23978910"

/translation="MPQASPRGYTLKVADAKVAKQIDAEAAAMELTPLAELMAAYGV
 RGVAVRKVLKEEGRAEVRGWRVVRDGEQDRRPLVERIADYFRVDGLKVGDDF
 PSTSLADRFGASRPTVAKALDRLETSLASASRQGRKRTVLALPQREQS"

gene complement (14012. .14485)

CDS complement (14012. .14485)

/gene="mutTSL"

/gene="mutTSL"

/codon_start=1

/transl_table=11

/product="mutT-like protein"

/protein_id="BAC21274.1"

/db_xref="GI:23978911"

/translation="MGRIDYLDHDPDAPPANSVVSVAVFQDQGRVLMIQRSNDRGW
 ALPGGGHDAGESISDVTVREVWEETGIDAEVIDSGIYTDPGHVMAYDDGEIRQQFSI
 CPEARPTGGEVRTSETTQVRWVAPADLVDELVHPTWRLRIEHMDRTRTPYIG"

ORIGIN

Alignment Scores:	1.56e-05	Length:	15397
Pred. No.:	673.50	Matches:	652
Score:	31.94%	Conservative:	295
Percent Similarity:	21.99%	Mismatches:	1156
Best Local Similarity:	5.10%	Indels:	872
Query Match:	1	Gaps:	122

US-09-522-753-5 (1-2517) x AB093554 (1-15397)

Qy	45	HisSerArgAspTyrAlaSerHisLeuSerProGlySerIleLeuGlnProGlnArg	64
Db	3815	CACAGCGGCACATCCCGCCGACAC-----CCGACAGCGCGGTATCCCCACATCAGCGC	3868
Qy	65	ArgProSerLeuSerGluPheGlnProGlyAsnGluArgSerGlnLeuHisLeu	84
Db	3869	GGGTGTGAATTAGGGCGGCACACCGGCCCGCCGACACCCCGCGCTACACCCCGCACACC	3928
Qy	85	ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGlu	104
Db	3929	AAACACGAGAGCGCGACCCCGGTTGCCCGC-----GATCGG	3964

105 SerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerProLeuLeu 124
| | | | | : : : : :
3965 CCGGAGCAGCGCGAG-GAACCCTGACCGAC-----CGCCCCAGTAT 4008
| | | | | : : : : :
125 AlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLys 144
| | | | | : : : : :
4009 CCCACCGGACCGCAGATCACCAGCGCCATCTG-----CTCGCCGACTA 4053
| | | | | : : : : :
145 LeuGluProValSerProProSerProHisThrAspProGluLeuLeuValPro 164
| | | | | : : : : :
4054 CTATGCCCGACACCGCGCGCGCGCGCGCGG-----CCCCCGCGCGCGCGCGCC 4110
| | | | | : : : : :
165 -----ProArgLeuSerLysGluLeuLeuLeuGlnAsnMetAspArgValAspArg 181
| | | | | : : : : :
4111 GGCCCGCGCGCGCGAGTTCAGCTCGCGCGCTCTTGGAG----- 4149
| | | | | : : : : :
182 GluLeuThrMetValGluGlnGlnIleSerLysLysLysLysGlnGlnLeuGlu 201
| | | | | : : : : :
4150 -----CTGGCCCG 4176
| | | | | : : : : :
202 -GluAlaAlaLysProProGlu--ProGluLysPro-ValSerProProIleG 220
| | | | | : : : : :
4177 GCCCGGTACCCAGACCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4236
| | | | | : : : : :
220 LuSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLysAlaGluA 240
| | | | | : : : : :
4237 GCTGTGCGCACCGC-----CGTCCGCGAGCGTCCGGATC 4269
| | | | | : : : : :
240 laAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnp 260
| | | | | : : : : :
4270 GGCCCGCAGCGCTTCTGCGGTGGCGCT-----CGAC 4305
| | | | | : : : : :
260 roSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysL 280
| | | | | : : : : :
4306 CTTCTCGCCACCG 4365
| | | | | : : : : :
280 euIleLeuTyrPheLysArgArgAsnHisAlaLysLysGlnTyrLysGlnLysPheCysG 300
| | | | | : : : : :
4366 TG-----CATCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4404
| | | | | : : : : :
300 lnArgTyrAspGlnMetAlaLeuGluLysLysValGluArgIleGluAsnAsp 320
| | | | | : : : : :
4405 ATCGCCC-GGTACGCTCCAAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4457
| | | | | : : : : :
320 roArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluI 340
| | | | | : : : : :
4458 CCAGACCATCACCAGTCGGTCAAGGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4517
| | | | | : : : : :
340 leArgLysGlnArg-----GluLeuGlnGluArgMetGlnSerA 353
| | | | | : : : : :
4518 GTCAGATCGACCGCGCGGAGCAGGTTCGTCGGCTCGTCTGGTGTATGCTCTTCGGCGCC 4577
| | | | | : : : : :
353 rgValGlyGlnArgLysGlySerMetSerAlaAlaArgSerGluHisGluValS 373
| | | | | : : : : :
4578 GTGCTCTTCAGCGTCTGACGGT-----CACTCCGCTCGTCGAGCGGTCACCCCGGAG 4631
| | | | | : : : : :
373 erGluIleIleAspGlyLeuSer----- 380
| | | | | : : : : :
4632 GGCTGGGCGTGGACAGCTGGATCCTGCCCTTCGTGTGTCAGCGCGCGCGCGCGCGCGCG 4691
| | | | | : : : : :
381 -----GluGlnGluAsnLeuLeuLysGlnMetArgGlnLeuAlaValIleP 396
| | | | | : : : : :
4692 GTCCGCTGCAGCAGATCACCAGCGTCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4751
| | | | | : : : : :
396 roProMetLeuTyrAspAlaAspGlnGlnArgIleLysPhe-----IleAsnMetAsnG 414
| | | | | : : : : :
4752 CTGCGTGGTCTACCGCGCGGAATGACCTCTCGGCTCAACATCGCGCGACTCGATCTCAAG 4811
| | | | | : : : : :
414 lyLeuMetAla-----AspProMetLysValTyr---L 424
| | | | | : : : : :
4812 GGCAGCTGGGTGGCGGTTCAGCGCGGTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4871
| | | | | : : : : :
4812 GGCAGCTGGGTGGCGGTTCAGCGCGGTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

424 ysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysP 444
| | | | | : : : : :
4872 GAGCGCCCATCGCATCGCGCGCGCATCACCGCGCGCGTCCGCGCATCGAGAG--- 4927
| | | | | : : : : :
444 heMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg-LysThrVal 463
| | | | | : : : : :
4928 -----TAGCCGGCGCGCGAGCGCGCGCGCGCGCGGTTCGAGCGGAGGACCGGTG 4979
| | | | | : : : : :
464 AlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuVal 483
| | | | | : : : : :
4980 AGCGCG-----CAGAGCGCAAGGCCCGGAGGAGCGCTCGGCGCGCGCGAG 5024
| | | | | : : : : :
484 ArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGln 503
| | | | | : : : : :
5025 CGTGAGCAGCAGCGCTGGCGCATGACGCGCGCGCGTGTAGCAGGAGGCGGACGCGTGA 5084
| | | | | : : : : :
504 GlnGlnGlnGlnGlnGlnProMetProArg-----SerSerGlnGlnGluLys 520
| | | | | : : : : :
5085 CACGCCGTGAGCAGGCGCGCGTGGAGCGTCTCGTCTACGACGCGCGCGAGCAG 5144
| | | | | : : : : :
521 AspGluLysGluLysGluLysGluLysGluLysGluLysProGlu----- 537
| | | | | : : : : :
5145 GCCCG 5204
| | | | | : : : : :
538 -----ValGluAsnAspLysGluAspLeuLysLysLysLysLysLysLysLys 553
| | | | | : : : : :
5205 CGTCAGCGCGCGCGAGGAGGAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5264
| | | | | : : : : :
554 SerGlyGluAspAsnAspGluLysGluAlaVal-----AlaSerLysGlyArgLysThr 571
| | | | | : : : : :
5265 GCCAGCGCGCGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5324
| | | | | : : : : :
572 AlaAsnSerGlnGlyArgArg-----LysGlyArg 581
| | | | | : : : : :
5325 GCTGAGCG 5384
| | | | | : : : : :
582 IleThrArgSerMetAlaAsnGluAlaAsnSerGluAlaIleThrProGlnGlnSer 601
| | | | | : : : : :
5385 GTGAGCACCACCGTGTAGCACCCTCGCTCAC---GACACCGAAGACCCCTCCGAGACC 5441
| | | | | : : : : :
602 AlaGluLeuAlaSerMetGluLeu----- 609
| | | | | : : : : :
5442 CCCAAGCAGCG 5501
| | | | | : : : : :
610 -----Asn 610
| | | | | : : : : :
5502 CGCCCGCGGTGAGCACACCGGTGAGCACCCCGCTCACGACACACAGAACCCCTGCG 5561
| | | | | : : : : :
611 GluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuGluHis 630
| | | | | : : : : :
5562 GAAACGCTCAAGATGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5621
| | | | | : : : : :
631 GlyArgAsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLys 650
| | | | | : : : : :
5622 GGCGCGTTCACAGAGCGAGGTTCGCGCGCTCACGCGCTGGTCCGCTGGTGGTGGCGCG 5681
| | | | | : : : : :
651 AsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLys 670
| | | | | : : : : :
5682 CGCTTC-----CAGGACCTTGACGACGCGGG 5708
| | | | | : : : : :
671 LeuLysMetGlu-----LysGluArgAsnAlaArg 681
| | | | | : : : : :
5709 GTGAGCGCGCGTGTGAGCACCGACCTGAACCGCATGACCGCGCGCGGTCCGCGCGCG 5768
| | | | | : : : : :
682 LysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProValValGlu 701
| | | | | : : : : :
5769 TCACCGCGCACCGACACCAAGCGCGGACTCTCTCATCGGTTCGTCGGCGCGCG--- 5819
| | | | | : : : : :
702 AspGluMetGluAlaSerGlyValSerGlyAsnGluGlu-----Met 717
| | | | | : : : : :
5820 -----TCCTGGCGCGCGCGGAGCATCCTCCACGACGCTCCCGATGACCTCCCGCGTT 5873
| | | | | : : : : :
718 ValGluGluAlaGluAlaAlaSerGlyAsnGluValProArgGlyGluCys--- 736
| | | | | : : : : :

Db 5874 ACATCGGCGGCGCATCGGCGCGCTCTGG---CCGCGCGCGCGCTCTCTCTGCTCC 5930
QY 737 -----SerGlyProAlaThrValAenAenSerSerAspThrGluSerlleProSerPro 754
Db 5931 ACGTGGTCCGCGCGCGCTCGGCGCGGACACGCGCTGGCGCGCTGCTGGCGCGACCTCACCC 5990
QY 755 His-ThrGluAlaAlaLysAspThrGlyGlnAenGlyProLys-----Pr 769
Db 5991 CGCAACAGCTCTTGGCGCGGACCGCGGGAACCTGGCGCGGACGTCGTCTGGCGCTCTCC 6050
QY 769 oPro-----AlaThrLeuGlyAlaAspGlyProProProGlyProPr 783
Db 6051 GCTCGCGCGCGGAGATTACCGGCTGCGCCACGAGATCGACTCTCA-CGCGC 6109
QY 783 oThrProPro-ArgArgThrSerArg----- 791
Db 6110 GCGCGCGCTCTGTCGCGCTGCTCTCGCGCTGCGCTCGAGTGGCGATGACCA 6169
QY 792 -----Alap 793
Db 6170 CCCTCATCCGATCGTGTCTGTA CGGCGTGGCCACGACGACGAGCGCGGCCCGCC 6229
QY 793 rolleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaProp 813
Db 6230 CTTCTGCTCGACACACACGAGTGGCGAACCGCCCGACGACCGCGCGCTCGCGCGCG 6289
QY 813 roSerProSerAla-----ProProProv 821
Db 6290 CCTCACCGAGTACACCGGCTCGACCCCGAAGTCCACCACTAGTGTATGATGACCAACCCCGG 6349
QY 821 al-----ValProLysGluLul 827
Db 6350 CGCCCGCCAGCTGTCTGCCCGCCACGTCGCGAGATCGATGTCGCGCGCGAGCGCGCC 6409
QY 827 ysGluGluGluThrAlaAlaProProValGluGluGlyGluGluGlnLysProProA 847
Db 6410 GACCGCGCTCGAGCT 6469
QY 847 laAlaGluGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCyst 867
Db 6470 GCACAGCTCGCGCGCGAATTTGCGCGCTCGACCCACGACCTCCAGC-TCCACACCGCGCA 6528
QY 867 hrGluGluAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaThrA 887
Db 6529 CATCAACCGCGCGCT 6588
QY 887 laGluGlyAlaLeuLysAlaGluLysLysGlu-GlyGlySerGlyArgAlaThrAla 906
Db 6589 AAGAAAGACGGCTGCGCGACACCGGCGCGAAGCGGTACACGGCGAGCGCGCTCGCGG 6648
QY 907 LysSerSerGly-----AlaProGlnAspSerAspSerAlaThr 920
Db 6649 GCCAGGAACGGCGCGGCAACAAACCCAGAGCGCGGACCCCGAGTACCACCGCGCAAC 6708
QY 921 CysSerAlaAspGluValAspGluAlaGluGlyGlyAspLys-----AsnArg 936
Db 6709 CCCACGCGCCCAAGCTCTGACCGCGCATGGACACGATCCCTGTCCTGGACTGGGCGA 6768
QY 937 LeuLeuSer-----ProArgProSerLeuLeu-----ThrPro 947
Db 6769 CGCTCTGACCTTCGACCCACCGGACCGGCGCTCTGCTGTGCGGCGCGCGCGCC 6828
QY 948 ThrGlyAspProArgAlaAenAlaSerProGlnLysProLeuAspLeuLysGlnLys 967
Db 6829 GATCGGTCCACCGCGGAGAGCCGCTCCAA-----GGTCTGCGCGCGAGCAGTG 6879
QY 968 Gln-----ArgAlaAlaAlaProPro----- 975
Db 6880 GAACCGCATCACCGGACCGCGCGCTCTACCGCGCGCGCGGACAAAGGGGACCGCGAC 6939
QY 976 -----ileGlnValThrLysValHisGluProProArgGluAspAlaAla 990

Db 6940 CCGGCGCGCGGACATCGGCGACACCGCGTTTCCAC-----TCCGACGCGCCA 6987
QY 991 ProThrLysProAlaPro-ProAlaProProProGlnAenLeuGlnProGluSerAs 1010
Db 6988 GCACCAACCTTCAGCGCGCGCGCTGTTGCGCGCTGACCCCAAGAAACCCGCTCGAC 7047
QY 1010 palProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaPro----- 1028
Db 7048 CCGCATCCACTACGCGACTGCGCGCGCGCGCTGATCGTGA-CCGACACCCCTGAC 7106
QY 1029 -ProAlaAspLysGluAla-----PheAlaAlaGluAlaGlnLysLe 1042
Db 7107 CCGCGTCCCGGACTGCGCGGCGAGCGGCGGTGAACCGGACTACGCGACACGAG 7166
QY 1042 uProGlyAspProCysTriThrSerGlyLeuProPheProValProProArgGluVa 1062
Db 7167 ACCGCGAGTACGCGCGCA---CCGAGTGGGACCCCTGCACCTGCTGGGACGAGAACC-- 7221
QY 1062 lleLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHi 1082
Db 7222 -----GCTCTGGAGCTCC-----TCCGCTCCCGCGCA 7250
QY 1082 sProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgPro----- 1099
Db 7251 TCCGCGCGCG-----CCGCGAGCGGTACAGCGACCCGCTGGGCGCC 7289
QY 1100 ---ProThrIle-SerAsnProProProLeuIleSerSerAlaLysHis----- 1114
Db 7290 AGCACCAAGCTACAGCGACGACCGCTTCTAACCCCGAGAGACAGCTTGTTCG 7349
QY 1115 -----ProSerValLeuGluArgGlnlleGlyAlaIleSerGlnG 1128
Db 7350 AGATCCGATCATCTCGGACCCCGCGCACCGCGCTCCCGCGCGCTCGCG-- 7405
QY 1128 lyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProV 1148
Db 7406 -----GGCACCC 7412
QY 1148 alThrMetGlyLeu-ProLeuProMetAspProLysLysLeuAlaProPheSerGlyVal 1167
Db 7413 TCACCTCTCGGACGCGCGGAGTACCCACCGCGAGTACGCGACGACGACGCTGTACA 7472
QY 1168 LysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValPro 1187
Db 7473 TCACCGCGCGACCTGCCACCCCGGAGGACTGCCCGCGCGCGCGCG-- 7526
QY 1188 ThrAlaGln---GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGly 1206
Db 7527 ACCTCGAGGCGCGCGCTCTGTCACGAGA-----TCAGTTGGGTGCG---CCCGGACCG 7577
QY 1207 SerIleThr-----LysGlyIleProSerThrArgVal 1217
Db 7578 TCACCGAGCGCTTCGCGGAGAGCGCGCGGAGTCTCTCTCGCAAGCGCGACT 7637
QY 1218 ProSerAspSer-AlaIleThrTyrArgGlySerlleThrHisGlyThrProAlaAspVa 1237
Db 7638 CTTGAGCGGATCGCTCGATGTCGACGCGCGCGCGCGCGCGCGCGCGAGCGAGGC 7697
QY 1237 lleuTyrLysGlyThrIleThrArgllelleGlyLysAspSerProSerArgLeuAspAr 1257
Db 7698 CGCGGAGAGCGCGCT---CACGAGCTCTGCGAGTTTCGACCTCAACGCGGCGGCGGACCA 7754
QY 1257 gLyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysHisVa 1277
Db 7755 CCAGCGCGCGCTTCTGCGCGCGAGCACCCCGCTCCAGCGCGCGCGCGCGCGCTACGT 7814
QY 1277 lleuSerTyrGluGlyMetSerValThrGlnCysSerLysGluAspGlyArgSerSe 1297
Db 7815 CCGCGAGGATACCGCGACTGGACACGACCCACCGGTAGCAACGCGCGCGCGCGCGCC 7874
QY 1297 rSer-----GlyProProHisGluThrAlaAlaProLysArgThrTyrAspMe 1313
Db 7875 ATCCGACCAAGAACCGCGCGCGCTCTGCTGCGACGACCCCTCAAGAGAACTGAGAC-- 7932

Qy 1313 tMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyAr 1333
Db 7932 ----- 7932
Qy 1333 gAlaIleProGluArgHisSerProHisHisLeuLysGluGlnHisIleArgG 1353
Db 7933 -----ACCCAGCATGA-CCCAACCGA-----CCCTC 7957
Qy 1353 ySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgAr 1373
Db 7958 ATCCGGCGAGAGCGGCCCCCGC----- 7982
Qy 1373 gGluAlaLysLeuLeuLysArgGluGlyThrProProProProProSerArgAspLe 1393
Db 7983 -----CGACGCTGTCGCCGCCCTCCAGCGCGCGAGCGGCTGG 8023
Qy 1393 uThrGluAlaTyrLysThrGlnAlaLeuGlyProLeu-----LysLeuLysPr 1409
Db 8024 CACGTCTTCCCACTCCCGCCGGGTACGAAGCGGCTCTCCCTGCACGGCGAGCGCTGCC 8083
Qy 1409 oAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIlePr 1429
Db 8084 CCGCACCGCCACCTCGTGTGACGACACGAGTGAAGTGGAGACCGCGGCCACACCGACC 8143
Qy 1429 oArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluG 1449
Db 8144 GCGAGCGGAT-----CGCGG 8158
Qy 1449 ySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
Db 8159 GCGCGTGGGGCGAGG---CCCGTTCAACGTCGCTCTGGCCACCGGGCGCTCGGGGCTCG 8215
Qy 1469 sLysHisAspValArgSerLeuIleGlySer-----ProGlyArgThrPheProPr 1486
Db 8216 TCGTGTGTCACCTCGACATGCCCAAGAGGGCGCCAGCGGCTCAGCGGTGGAGGTGT 8275
Qy 1486 oValHisProLeu-----AspValMetAlaAspAlaAlaLeuGluAr 1501
Db 8276 TCGCGGCCCTCTGTAGCGCGCGCGCCAGCGCTTCCTCGCGACCTTCCTCGGCGAGGACG 8335
Qy 1501 gAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySe 1521
Db 8336 CGTGGGTGGCGAGCACCTCTACTTACCGCCCGCGTGC-GCATCCGCTCGGCACTCC 8394
Qy 1521 rIle-----AlaArgGlyAla-----Pr 1527
Db 8395 GCTGGATTGCTAGGCCAGTTGGTGGACACGCGGGCGTGGGGCGGCTACGTCGTCACCC 8454
Qy 1527 oValIleValProGluLeuGly----- 1534
Db 8455 GGGAGCATCACCCCGCGGGGTGTACACGGTCTCTGGACGACGCTCCGCTCCTTCCTCTG 8514
Qy 1535 -----LysProArgGlnSerProLeuThrTyrGluAspHi 1546
Db 8515 CCCAGCTGGCTCTTTGAGGCACTCACGCGCGCGCCAGAACTCT-----CA 8559
Qy 1546 sGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProTh 1566
Db 8560 CGGC---CCCGTGAGGGGTTCGGGCTTCCCGCGGTCTAGTGGCAGCCGCGCGGCGCGT 8616
Qy 1566 rProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuTh 1586
Db 8617 GCGGCG-----TGGAGCGCGAGTGGCGCAACGTCACGGCGCGCCGACG 8661
Qy 1586 rSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisProHi 1606
Db 8662 AAGCAGGCCAACACACGCTCAACCGGTTCGTCATTCCTG-GTGGG----- 8706
Qy 1606 sProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSe 1626
Db 8707 -----CGCTTGTGTCATGGGGCGACATCGACCGCA 8738

Qy 1626 rHisIleProLeuAlaPheAspProThrSerIleProArgGlyIle-ProLeuAspAlaA 1646
Db 8739 GGTGTGGAGGAGCATTCAGGGGGGGGAATCGGGGACTCACCGCGCGCGAGTG 8798
Qy 1646 laAlaIaTyrTyrLeuProArgHisLeuAla-----ProA 1658
Db 8799 CCGCG-----CCACGATCCGTAGCGCCTCGACAGCTCCCTCCCAAGGCCG 8846
Qy 1658 snProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaA 1678
Db 8847 GCCCGGGAGGTGGCATGAGGCCCAAGATCCCTCCCTTGGATCGCTCC----- 8899
Qy 1678 laLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMethHisAla 1698
Db 8900 -----CACCAAC 8906
Qy 1698 snThrAla-----ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerP 1715
Db 8907 ACCCGGCCCTCGCGCGCGAGCGGTAGGACTGCGGGCCGTGGCGGCCGCGAAAGGTC 8966
Qy 1715 roArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGly-----I 1731
Db 8967 GTCCCGAAGCGCTCTTCGGGTCTGACCCGACCCCGCGCGGGGTGAGGATCCTCGC 9026
Qy 1731 leIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGlyThrP 1751
Db 9027 GATGTCCACATCGGCGCGCCCGCGCAACGCGATGCCACCGCGCACTCTCTGTGTC 9086
Qy 1751 roAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerA 1771
Db 9087 CTGCGCGCGGACCTGAAGCGCTTCGGCCACACGACAGATCGCCC----- 9130
Qy 1771 rgHisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrT 1791
Db 9131 -----GCCTCGACCGCACACCCGCC-----ACCGCGCCAGTGCCTC 9170
Qy 1791 hrSerSerSerGluArgGluArgAspArg----- 1802
Db 9171 GCTGTCAACCAAGGAAGAGCGCGCATCGCGCACGACCCCGAGCTGTGGCGCGG 9230
Qy 1803 --GluArgAspArgAspArgGluArgGlnLysSerIleLeuThrSerThrThrThrValG 1822
Db 9231 TTCGACGAGATGACGCGCGAGGATGTCGCGGGCCCATCTCTGGGACGACCGCTACCGCTC 9290
Qy 1822 luHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySer----- 1837
Db 9291 AACCGCGCGGCA---GCTCCCGTGTTCCTCGTGACGCGCTGCCCGACTGCTCGCC 9347
Qy 1838 -----SerGlySerSerGlyGlyGly-----GlySerSerSerA 1850
Db 9348 GCGATGACGCGCGGGTGGCGGAGGAGACCCAGACCCCGCTCGACCTGGCGGGGTGCTC 9407
Qy 1850 rgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArgThrGlnAspA 1870
Db 9408 GCCTCTCGCGCTCTCGGCACCGCGCGCGGTGCGCTCACCGTACCGCTGCGCGG---G 9464
Qy 1870 laLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaV 1890
Db 9465 CAGTGAGCGAGCGGTCAACCTCTACACCGCGT----- 9499
Qy 1890 alGluProSerLysProThrValLeuArgSerThrSerThrSerSer---ProValArgP 1909
Db 9500 --CGCGC-TCCCGCGCGCAACCGCAAGAGCGGTCTTCGGGTCTCATGACCAATCGCT 9556
Qy 1909 roAlaIaThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValT 1929
Db 9557 CCTCG-----CGCGCGAAGGCGCTCATCGAACTCACCGCGCGCGCGCGCGCA 9607
Qy 1929 yrProThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProG 1949
Db 9608 GCGCGCG-----CCTCGCGCGGATCGCGAAGCGCGCGCGCA 9646
Qy 1949 luArgProArg----- 1952

Db 9647 GAAGCCGAGAGCTGGCCGCCAAGCGGAGCGGCGAAGCTGGCCGATCTGACCGCCA 9706
QY 1952 -----
Db 9707 GGCGGTGTCCTGTCGAGGCGCGAGCGGCGCGTGTATCCCGCGGAGCGCGAGCTCGT 9766
QY 1953 -----AlaAspThrGlyHisA 1958
Db 9767 CGCGGACGAGTACCGCAGAGAGCTGACACCTCTAGCCAGCAGGACGGCGGAT 9826
QY 1958 laPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysG 1978
Db 9827 CTCGATCTGTCCTCCCGGAGGAGATCT-----TCGAGATCATCGCAGG 9871
QY 1978 lYserGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrp 1998
Db 9872 CCGTACTCGGGGCGCCGAAACATGGGATCTTCTCAAGGGCCACGCCGCGACATGGC 9931
QY 1998 roAlaLysAsnLeuAlaPro-----HisHisAlaSerProAspProProAlaProProA 2016
Db 9932 CCGGTCAACCGTCAGCGCCGAGCGGCGGAGCGGCGAGTACATCGAAGACCCGCCATACCATGGG 9991
QY 2016 laSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluL 2036
Db 9992 CTTGGCGTGCAGCCG-----ACGTCCTGGACGCCATCGGC----- 10029
QY 2036 euGluLeuArgSerLeuGlyTyHisGlySerSerTySerProGluGlyValGluProV 2056
Db 10030 -----AGATCAAGGGCGCCGACGCGCGCGCTCTCGCCGCTTCTCTACTCCCG 10081
QY 2056 alSerProValSerSerProSerLeuThr-HisAspLys---GlyLeuProLysHisLeu 2074
Db 10082 GCCGAGTCCCTGTCGTACCGCAACCTCACCCCGAATCTCTAGCCCGGACAT--- 10138
QY 2075 GluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProVal 2094
Db 10139 -----CGCCGACCTACGCCAGAGCTCGGCGCTAGCCCTGACCCCTCGC 10186
QY 2095 LysLeuGlyGlyGluAlaHisLeuProHisLeuArgProLeuProGluSerGlnPro 2114
Db 10187 GGGCTGACGAGAGCCCGCAACTCACCTCAGCCCGA-----GGCCGACGCCGCTCT 10240
QY 2115 SerSerSerProLeuLeuGlnThrAlaProGlyValLysGly----- 2128
Db 10241 CTTGGCCTACAGCGGTACCGAATCCCGCTCCGCAAGAGCGGGCCCTCGGCGCGAT 10300
QY 2129 -----HisGlnArg-ValValThrLe 2135
Db 10301 CATCAACTGGGCCAGCAAGCGGAGCGGCGCGTCCGCGCATCGCGGACTCTCCACCT 10360
QY 2135 uAlaGlnHisIleSerGluValIleThrGln----- 2145
Db 10361 CGCCAAGCATCCCGAGAGCGGTGGACCGCGCGCATCGCGCGACACGATGGCGCGCGC 10420
QY 2146 -----AspTyThrArgHisHis-ProGlnGlnLeuSerAlaPro---LeuP 2160
Db 10421 CACCGGCTCGGGACTACTTACCGCCCGAGCCCTCGAGTCTTCAACCCCATGAAGGC 10480
QY 2160 roAlaProLeuTySerPheProGlyAlaSerCysProValLeuAspLeuArgProp 2180
Db 10481 TGACCCCGCTCAGAGGCGCGCCGACCCCTCTGACGACCTGACCGAGACCCGACCGC 10540
QY 2180 roSerAspLeuTyLeu-ProProProAspHisGlyAlaProAlaArgGlySerProHis 2199
Db 10541 CGCCTTCAACAGCGGTGACCTCTTCGTGTCAT---GCCCGCGAGCGA---GTTCCCGCAC 10594
QY 2200 SerGluGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlu 2219
Db 10595 CATGAGGACCTGACCCCGCCCTGAACTCTCTGGAGGACACCGCTGGGT----- 10645
QY 2220 AspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArg--- 2238

Db 10646 -----CGCCAGCAGCCCGCCCGCCCGCCAGCACCACACACGCGCGG 10684
QY 2239 ---SerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMet 2257
Db 10685 CGGCCACCTCTCCCGCTA-----CGAGACCCACCCCGCATCACCCCGCGCGCTG 10738
QY 2258 GlySerLysSer---ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThr 2276
Db 10739 AAACCCCGCCAGCGCGCGAGCAGAACTGACAGAAACCCCGCAGCCATCTGACCTGCGA 10798
QY 2277 Glu----- 2277
Db 10799 AAACAAGCGCTTGGCGCTCTGACAAAACCTCCGGAACCTCCGACAGAACTCAGAGAAC 10858
QY 2278 -----SerAsnSerAlaMetValLysSerLysLysGlnGlu 2289
Db 10859 CTCACGCGCAGCGGAGCAGAACCTCCGAGCGCGCCAGGTTCGAGGTTCTGTCCAGAG 10918
QY 2290 IleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyraAsnIleSerGlnPro 2309
Db 10919 TTATTCGAGGTTTG---TCACGACAGCCCATCAGCCGATCGGAGGTCCAGCGG-CCA 10974
QY 2310 GlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyArgSer 2329
Db 10975 GACGCGGAGGTCTGTGAGTTCTGTGAGCACTC---CCAGGGGTTCGACACCCCGCGCC 11031
QY 2330 GlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleArg-LysAl 2349
Db 11032 GAGCTGAAGAGAGCCCATGGCGAA-----GGCCAAAGACGCCCGCGCTCAACAAC 11082
QY 2349 aLeuMetGlyLysTyAspGlnTrpGluGln-SerProProLeuSerAlaAsnAlaPheA 2369
Db 11083 CTTGCCGAGCGATACCTCACCCCGAGACATCGCGCTCTCTCGAGGTCCCGCTGGAG 11142
QY 2369 sn-----ProLeuAsnAlaSerAlaSerLeuProAlaMetProIle- 2383
Db 11143 ACCGTCTACCACTGGCGAAGCAGCACACCGCGCGCGCGATTCGTTGCGCGGCCAT 11202
QY 2384 -----ThrAlaAlaAspGlyArg--- 2389
Db 11203 GTCCGCTACGACCCCAACCGTCCAGCGGTGGTGCACAGCAGACACAGGTCCGACGCC 11262
QY 2390 -----SerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysVal 2405
Db 11263 GCCGCTGACCCCGCAGCACCGCTGCTGCTCCCGCTCCAGTGGCGGAGCGGCGCT 11322
QY 2406 SerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAsp 2425
Db 11323 TCCGCTACTCCGAGAGGATCCACAGCACCATGGCGGCCACGTTCAAGACCGCTGTATCA 11382
QY 2426 ArgProProSer-----ValSerSerValHisSerGluGlyAspCysAsnArgArg 2442
Db 11383 AGACCGAACTCGGCTCGAGCGGCAAGCTCTGTCAGTCA-----AGACGG 11427
QY 2443 Thr-----ProLeuThrAsnArgValTrp----- 2450
Db 11428 ACCGTTTCGCACTGGCACCGGCTACCGCGCCCGCTACGTCGCGCCCGCAGCGCACAGA 11487
QY 2451 -----GluAspArgProSerSerAlaGlySerThrpPro 2461
Db 11488 AGTCCAAGAGCTTCCCGCAGCAGCAAAAGCGACTCGCCGAGCAGTGGCTCAACGAGACCG 11547
QY 2462 PheProTyraAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSer---ProPro 2480
Db 11548 CGCCGACATGGCGCA-----GGGTCAATACATCGACCCCGCG 11586
QY 2481 ProProGlyLeuProAlaGlySerGlyProLeuAlaGly 2493
Db 11587 CGCCCGGATCACCTTCGCGCAGTTCGCTCAGAGAAGTGGG 11625

RESULT 85
AX600586/c
LOCUS

75236 bp DNA linear PAT 14-FEB-2003

Db	55458	GTCCGCTCCACAGCATCCACATCGGACACGACCAACCCCGCACTCGTAGTGTGGT	55399
Qy	515	--SerSerGlnGluGluLysAspGluLysGluLysGluLysGluLysGluGluG	534
Db	55398	GATCACCTTTGTGGAGGACCATTCGCGCGGTCAGTCCTGTTGACCGCCCTCTG	55339
Qy	534	lulysProGluValGluAenAspLysGluAenAspLysGluAenAspLysGluLysThrs	554
Db	55338	ATTACCGCCGACCCAGAA-----CCACCGCAACACCCGAT--G	55300
Qy	554	erGlyGluAenAspGluLysGluAenAspGluLysGluAenAspGluLysGluAen	574
Db	55299	CCATTACGAGCGGATCCACAACTCTCCACAGCACTAGCCACCTCGCCCA	55240
Qy	574	erGlnGlyArg--ArgLysGlyArgLysGlyArgSerMetAlaAenGluAenSerG	593
Db	55239	GCCGTCCTCGCGCCCTCCGCGGACGACTTGCAGCGCCCATCGGTGCGACCCCG	55180
Qy	593	luGluAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAenGluSerS	613
Db	55179	CTGACGGAGACTCCAAACGTCGCG-----55151	
Qy	613	erArgTrpThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgA	633
Db	55150	-----G	55150
Qy	633	snTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPhet	653
Db	55149	GGTGCCATCATCGTCACG-----55130	
Qy	653	yrPheAsnTyrlsLysArgGlnAsnLeuAspGluLleLeuGlnGlnHisLysLeuLysM	673
Db	55129	-----CACCGCAAGGCGAGAT-----55112	
Qy	673	etGluLysGluAenAlaArgLysLysLysAlaProAlaAlaAlaSerGluG	693
Db	55111	--CACACTACCGGAACGAGTATTGACCTGCAGGTGCAACCCACGAGATGACGA	55054
Qy	693	luAlaAlaPheProValValGluAspGluMetGluAlaSerGlyValSerGlyA	713
Db	55053	ACACCGCGTGTCCACCGTCACCGGACGACCTCGAAACCGAACGATACCGCCGCGC-	54995
Qy	713	snGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProA	733
Db	54994	-----CGA	54991
Qy	733	rgGlyGluCysSer-----GlyP	739
Db	54990	GGCCACACTGCCCGCTGCCATTACCGAGATAGCTTCGAAGCCCTCCGCACTCCGCG	54931
Qy	739	roAlaThrValAsnAsnSerSerAspThr-----GluSerIleProSerProHist	756
Db	54930	CGGAACCGTCCGCGGTAGTGTGTATCATCAGCCCGCGGAACACACCCGCTCCGCTCC	54871
Qy	756	hrGluAlaLysAspThrGlyGlnAsnGlyProLysProAlaThrLeuGly----	774
Db	54870	GCGCAGCGACCCCGGATCGATCCAGCGGTTCAGCGCTCCAGGACAGTCCAGCAG	54811
Qy	775	-----AlaAspGlyProProGlyPro-----782	
Db	54810	CAACCGTTGTCGGGTTCATCGCCAGTGCCTCAGCGAGAAATCCCGAANAACGCGCG	54751
Qy	783	--ProThrProArgArgThrSerArgAlaPro-----793	
Db	54750	ATCGAACTCCGCGGTCCCGAGGAACCGCCCTCATCGACATACGAGTCCGGAACG	54691
Qy	794	-----IleGluProThrProAlaSerGluAlaThrGlyAlaProT	807
Db	54690	ATCGGATCTGGATCGAACAATCCGCGAGATCCAGCGCGATCGTCGCGGAGTCCCG	54631
Qy	807	hrProProAlaProProSerProSerAlaProProProValProLysGluGluL	827
Db	54630	GACCGCATCCACGCCCCCGGCCCAAAAGCTCTCCGGTGAGGACACACCCACC	54571
Qy	827	ysGluGluGluThrAlaAla-----833	
Db	54570	CGGATACGACACCCATCCGACTTCGGGACGAGATCGTCTTCGACGCTCGCGTCT	54511
Qy	834	--AlaProProValGluGluGluGluGluGluGluGluGluGluGluGluGluG	853
Db	54510	GTCGTCACCGCGCGCGCTCGTCCGACCTCGCTCCGCTGCT-----54467	
Qy	853	alAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluGluAlaGluG	873
Db	54466	-----CCTGGTCAGGAACCGCGCCACCGCGCGCTCGGTAGTCGATACGCGGT	54412
Qy	873	lyProAlaLysGly-----LysAspAlaGluAlaAlaGluAlaThrAlaGluG	889
Db	54411	TCCCGGACGCGCGCTCCAGCACCCCGCGCGCTTACGCAAC--TCCACACCAACA	54353
Qy	889	lyAlaLeuLysAlaGluLysGluGlySerGlyArgAlaThrAlaLysSerS	909
Db	54352	ACGAATCAAGCCCGA-----TCCCGAAAGCAGCTCGACCCCGACCT	54308
Qy	909	erGlyAlaProGlnAspSerSerAlaThrCysSerAlaAspGluValAspGluA	929
Db	54307	CGGCGCACCCGAAATACCCAGCACCTCGCGGTGCTCCGACACCCCAAGCAAGG	54248
Qy	929	laGluGlyGlyAspLysAsn-----ArgLeuLeu-----Ser-Pro	940
Db	54247	CGTCTCTCGCGAGGTGCGTGCCTACTGCGCTGATAGAGTGATGTAGTACCA	54188
Qy	941	ArgProSer-----LeuLeuThrProThrGlyAspProArgAlaAsn-----	954
Db	54187	GTGCGGTGAGCATCGCGGTAGCGTTCGCGCGCGCTGCTCGCGCAGGAGGCCCTGT	54128
Qy	955	---AlaSerProGlnLysProLeu---AspLeuLysGlnLysGlnArgAlaAla	972
Db	54127	CGAACCTCCCGGAACCGCTGCGCGATCCACGCGCGCGCATCGAACACCGCA	54068
Qy	973	ileProIleGlnValThr-----LysValHisGlu	983
Db	54067	AACCATCTTGTGGACACGCTGGACTGGAGCGCGCATCTGCTCGGTGACATCAC	54008
Qy	984	ProProArgGluAspAlaAlaProThrLysProAlaProProAlaProProGln	1003
Db	54007	TGAGCGCGCGGTATCCCGCTCGACTGCTCCACAGCCCGCCACCAACGACGCGAG	53948
Qy	1004	AsnLeu---GlnProGluSerAspAlaProGlnGlnProGlySer-----	1017
Db	53947	GCAATCCCTGCGCGCGGTATGCGCAATGCGCTCCAGGAACGTGTTCCGCGCGCAT	53888
Qy	1018	-----SerProArgGlyLysSerArg-----1024	
Db	53887	AATTACCTGACCGCGCGCGGTCAACCCCGACAGCGAGGAGAACACGCAAAACG	53828
Qy	1025	-----SerProAlaProProAlaAspLys	1032
Db	53827	ACAAGTCCATATCCCGAGTCAGCTCGTGCAGGTTCACGCGCGCGCTTCGGACCA	53768
Qy	1033	GluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProCysTrpThrSerGly	1052
Db	53767	ACACCG-----CCCCCAGCTCTCCGCTGTCAGCATCGA	53732
Qy	1053	LeuPro---PheProValProProArgGluValIleLysAlaSerProHisAlaProA	1071
Db	53731	TCACCCGTCAGCGACACACCCCGG-----CATGCACCAAC	53693
Qy	1072	ProSerAlaPheSerTyAlaProProGlyHisPro-----LeuProLeuGlyLeuHis	1089
Db	53692	CACACAGCG-----GAAACTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	53639
Qy	1090	AspThrAlaArgProValLeuProArgProProThrIleSerAsnPro-----Pro	1106
Db	53638	GATCCGCAACGTCCGACCGCCACCATCGACCCCGCACCGCAACCCCTCCAGTCTCA	53579

```
Qy 1107 ProLeuSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSer 1126
|||
53578 CCAACTCGGTGATTCGGGAGCATCCAGC-----CAC 53546
Qy 1127 GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGly 1146
|||
53545 GACGGCCCGTCA-----ACACCACTGACGAACACCGCGTCGCGCCACACAGGT 53498
Qy 1147 ProValThrMetGlyLeuPro-----LeuProMetAspPro-----LysLysLeuAla 1162
|||
53497 -----GGCGTCCACACCGCCACCCACACAGCCGTCGCCACCGGAATCAGCA 53450
Qy 1163 ProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGlu 1182
|||
53449 CCG---TCCGGT-----CGGACACCGCCACACCGCGGCACCTCGACCGCGCCGGAA 53399
Qy 1183 SerLeuGlyValProThrAlaGlnGluAla-----SerValLeuArgGlyThr 1198
|||
53398 -----CCGACGCCACACCGCGCACTGTGCACACACCCCTTACGCACCGCCCACT 53351
Qy 1199 AlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPro 1218
|||
53350 GCGGTTCATCGACCCACCGCACTCGGCAACCGCGGCCACGACTCGCGCGTACCGTCCA 53291
Qy 1219 SerAspSerAlaIleThrTyrArgGlySerIleThr----- 1230
|||
53290 CATCGACCAACAACCGCGCGGTCTCTCCGACTGCGTGAACGCAACAACCCCAACA 53231
Qy 1231 ----- 1232
|||
53230 ACGGCGCATCGCCAGCTTGTATCGCGCGTACCCGCGAGCGCTGTCACCGCCACACG 53171
Qy 1233 ThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerPro 1252
|||
53170 TCACCAACCAACAGCTGACCGCGCCACCGCTCTCTCG-----CCA 53129
Qy 1253 SerArgLeuAspArgGlyArgGluAspSerLeuPro----- 1264
|||
53128 ACCACGACTGACACCGCGCCAGCAT-----CTCCGGGTCTCCGCGAGTCCGCCACGC 53075
Qy 1265 ----LysGlyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGln 1283
|||
53074 CCGCGTCGGGATCCGCTGTTACCGCGCGCAATAAACACGACATCAGCAACACAG----- 53020
Qy 1283 yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProHisGln 1303
|||
53019 -----CGCACCGGAATCAGGTTCGCCGCA 52994
Qy 1303 uThrAlaAlaProLysArgThrTyrAsp-----MetMetGluGlyArgValGlyArg 1320
|||
52993 GCTCAGCCACACCGCAACGATCGACGGAACCTGGTGGAGCCCAATCAGACCGGCCA 52934
Qy 1320 gAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHis 1340
|||
52933 TGCCGCGACAGCATCACAAGCAACCGCCACCGACCCGATACCGTCGACCGGTGCGCCA 52874
Qy 1340 sSer---ProHisIleLysGlnHisIleArgGlySerIleThrGlnGlyIle 1359
|||
52873 CCGGTCCCAAGTCACTCGACAGCGC-----GTCACCGACACAGT 52832
Qy 1359 e-----ProArgSerTyrValGluAlaG 1367
|||
52831 CCGGATTCGCAACCGCACCGACACCGCGCGCAACCGGAGAGAATCGATCGAGCGCA 52772
Qy 1367 lnGlu-----AspTyrL 1371
|||
52771 GCAACCGCCCACTCATCATCAGCCACCGACACCGCATCGGTGCGGTAGATCCA 52712
Qy 1371 euArg-----ArgGluAlaLysLeuLeuLysArgGluG 1382
|||
52711 CCGGTGCGGTGATGTCCTCCACCTGACCGGAAAGTTCTACCCCGGTAAAGAGAACG 52652
Qy 1382 lyThrProProPro-----ProProSerArgAspLeuThrGluAlaTyrL 1398
```

Db 51702 GCCGACCCCTGCTGCGGAGCGCCCGAACCCTGTCGGGAAGCGGAGCGGCTCACGCAC 51643
Qy 1684 -----ThrIleAAspTyrIleThrSerGlnGlnMetHisHisAsnT 1699
Db 51642 CTGACGCACCCAAATACCTGGTGTAGCGAGCATGCCAGCGGTGTGAGAGCTCACCGCTCAA 51583
Qy 1699 hrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerS 1719
Db 51582 CGTCGACACACCGCGGAGGTGCG-----AACGCCGTGTGTCACGC----- 51542
Qy 1719 erLeuAlaLeuAsnTyr-----AlaAlaGlyProArgGlyIleIleAspL 1734
Db 51541 -----TGCCTGGCATATCTGTGAACCTCTGCAGCATCGGTCCATCCGATACGAATGAAA 51487
Qy 1734 eu-----LeuValProProThr-----SerGlnV 1737
Db 51486 CGCATGCGAAACCGCAACACCGCCACCGAACCCCTTGCCCATCAGCCCGCGGCAAT 51427
Qy 1737 alProHisLeuProVal-----LeuValProProThr----- 1747
Db 51426 GTCTTAAAGCACCTCCCGATCGCGGAGAGACACACCGACTCAGAAGCGTTGACCGCGGC 51367
Qy 1747 ----- 1747
Db 51366 GATCCCAACCCCTTACCCCGAGCGGCCAACACGCGCGAGCTGCTCTCGCCAGCAGC 51307
Qy 1748 -----ProGlyThrProAlaThrAla----- 1754
Db 51306 CACCGCCAAATGGCAGCGCATCCGGCAACCGCTGCATCAACCGGGCAGCACGCCGCCAC 51247
Qy 1755 -----MetAspArgLeuAlaTyrL 1761
Db 51246 CAACCGAGCGCATTCGGGAACAGACAAACACACCGCGCGCAACCGCGCCCAACTCAC 51187
Qy 1761 euProThrAlaProGlnProPheSerSerArgHisSerSerProLeuSer----- 1778
Db 51186 CACCGAATGCCCATCACACACAGCGCGAACACACCCACAGCGCCAGCAGCGCTCAGCAA 51127
Qy 1779 -----ProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArg 1797
Db 51126 ACCAACTGCAACGCGAACAACCCGACTCGGCCCAACAGTCCGATCCAGCAAGGACCC 51067
Qy 1797 luArgAspArg----- 1800
Db 51066 GTCGGAACCGACACACATCCCGAACCCCACTGCTGTCGCCAGATCGCATCCAGTCC 51007
Qy 1801 -----AspArgGluArg-AspArgAspArgGluArgGluLysSerIle 1814
Db 51006 GGCACAAGCTTCGTCAAAACGCTCAGCGAAACCGGAAACGCGGAGTAGAGCCCTCTTCC 50947
Qy 1815 -----LeuThrSerThrThrThrValGluHisAlaProIleTTPArgProGly 1830
Db 50946 CATCCCGACGCACTGACC-----ACCTGACCGGAACACAAACCC-----AACCCCGCC 50896
Qy 1831 ThrGlnGlnSerSer-----GlySer-SerGlySerS 1841
Db 50895 AGAGATCAGGATCCCGACACACCACCCAGAGCGTCGCGACCGCGCAGCGAGCTTCCAA 50836
Qy 1841 rGlyGlyGlyGlySerSerArgProAla-----SerHisSerHisAl 1857
Db 50835 CCCAGACAACAGCGCTTACCGTCCGCTCCAGCAGCAGCGCCGCTCTCTCCAAACGCGGC 50776
Qy 1857 aHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnGlnArgProSer-- 1876
Db 50775 CCGGGAAGAACCAACGAAACAGCGGCATCAAGGGATGAATATCGGGAGATTATCCAA 50716
Qy 1877 -----ValLeuHisAsnThr-- 1881
Db 50715 ATAGGACATCAATCGGCTCGCTGAGCGCTGAGCGCTTCGGTGTCTTCGCCGACAGAT 50656
Qy 1882 -----GlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu----- 1898
Db 50655 CAACGGGACCGCTGGGATGTCCACGGTAGATTCCGATTTCGGGACCGTGTGCTGCGCAAT 50596

1899 -----ArgSerThrSerThrSerSerProValArgPr 1909
Db 50595 CTCGCTCGTGTATTTCGAGGCTTGTTCGAGGATGACATGTGCTGGTGGCTCGCTGATGCC 50536
Qy 1909 o-AlaAlaThrPheProPro-----AlaThrHisCys----- 1919
Db 50535 GAACGACGACACTCCCGCCCGACGAGGATGGCGGTGTCGGGCCAGGGCGTGTCTCCGT 50476
Qy 1920 -----ProLeuGlyGlyThrLeuaspGlyValTyrP 1930
Db 50475 AAGGAGCTGGACCTCCCGCCCGCCGACCAATCCACTT-----C 50440
Qy 1930 roThrLeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluA 1950
Db 50439 CGCGTAGTGTTCATCCACATGCAACGTGGCGGGAGCTGCTGTCGCCGATCGCCATCAC 50380
Qy 1950 rgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuG 1970
Db 50379 CATCTTGATCACACAGCAACACCCGCTG-----CCGCTGCGTATGACCAATATT 50329
Qy 1970 luProAlaSerSerProSerLysGlySerGluProArgProLeu----- 1984
Db 50328 CGACTTCACCGACCCCAACCAACGCGCATCCCTATCCGCTCTGCTCCCGTAAGTGGC 50269
Qy 1985 -----ValPro--ProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnL 2002
Db 50268 AATCAGTGTCTGCGCTCAATCGATCACCAAGCTGTGTCCAGTCCCATGCGCTCCAC 50209
Qy 2002 euAlaProHisHisAlaSerProAspProProAlaPro-----ProAlas 2017
Db 50208 AGCGTCCACATCAGACACCGACAGACCCGCACTCGCAACGCGTGTGCTGATCACCCGCTG 50149
Qy 2017 erAlaSerAsp-----ProHisargGluLysThrGlnSerL 2029
Db 50148 CTGCGACGAGCAATTCGCGCGCGTCAACCATTCGACGCGCACCATCTGTGATTCACCGCGCT 50089
Qy 2029 yaPheSerIleGlnGluLeuLeuArgSerLeuGlyTyr----- 2043
Db 50088 ACCCGCAGCAGCCGACGAGAACCGCGTGGTGGCGCGGCATCCGACAGCGCTCCAA 50029
Qy 2044 -----HisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSer- 2061
Db 50028 CAACACGAGCGCGCTTTCGCCCGCGGTGCCATCCGACGAGCGCGCAACACACT 49969
Qy 2061 ----- 2061
Db 49968 ACACCGCGCTGCGCGCCCAACCCCTTGGCGCGGAAACTCCACGAACATCCCGCGCGT 49909
Qy 2062 --ProSerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSerH 2081
Db 49908 GGCCATCCCGTCCAC-----CACCGCAAGAGCCAGCATCACACTC 49867
Qy 2081 isLeuGluGlyGluLeuArgProLysGln----- 2090
Db 49866 ACCAGACCGAGTGTCTTGA-CCGCGCAGGTGCATGGACCAACAGACGACGAAACACGCG 49808
Qy 2091 -----ProGlyProValLysLeuGlyGlyGluAlaAlaHis-----L 2103
Db 49807 TATCCACCGTACCGGAGGACCTCTGAACCAACGAATACGCGACCCCGCAAGAAGA 49748
Qy 2103 euProHisLeuArgProGlu-----SerGlnProSerSerSerProLeuG 2121
Db 49747 CGCGCGCGCGCTGCCATTACCTAGATAACCTCGAAACCTCGCGCGCTGGTACGA 49688
Qy 2121 lnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAlaGlnHisIle---- 2139
Db 49687 ACCGCGCGCGTA-GTCGTGTGTGCATCAGCCC-----CGGAACACACCGCGTC 49641
Qy 2140 -----SerGluV 2142
Db 49640 CGACTCCTCGCAACGAAGCGGATCAATCCCGCGCGCTCCAGGCTCTCCCAAGCATC 49581

(100.0% identity in 361 aa overlap). Probable adhE2.
zinc-containing alcohol dehydrogenase, similar to several, especially mycothiol-dependent formaldehyde dehydrogenase from Amycolatopsis methanolica P80094 (360 aa). Contains P80059 Zinc-containing alcohol dehydrogenases signature. FASTA scores: >sp|P80094|FADH AMYME
NAD/MYCOTHIOL-DEPENDENT FORMALDEHYDE DEHYDROGENASE (MD-PALDH) length = 360, Expect = e-156, Identities = 268/358 (74%). Also similar to Rv0162c, (MTC128.02c, 35.0% identity in 371 aa overlap)."
/codon_start=1
/transl_table=11
/product="Probable zinc-dependent alcohol dehydrogenase AdhE2"
/protein_id="CAD97144.1"
/db_xref="GI:31619032"
/translation="MSQTVRGVIAKQGFVVELNIVVPPGPEAAVDVDTACGVCHT DLTVREGGINDEYFLLGHEAAGII EAVGPGVTA VEPFGDVIINWRAVGCQCRACKRG RPYCDPTNAEOKMTLTGCTELTAALGICAPADKTLVHSGCTKVDPAADPAVAGLL CGVMAGLNAINTGVTTRDDTVAVITGCGGVGDAIAGALVCAKEII IAVDITDDTKLD WARTGATHTVNAREVDVVOAIGLITDGFADVVIVDAVGRPTYQAQAFYARDLACTVV LVGVPDPELMDLVDFFSHGALKSSWYDCLPESDPTLIDILYLOGLRLPLQRFVVS ERIGLEDEAEAFHMGKGLRSVVM"
/locus_tag="Mb2284"
/locus_tag="Mb2284"
/note="Mb2284", len: 211 aa. Equivalent to Rv2260, len: 211 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 211 aa overlap). Conserved hypothetical protein, similar to hypothetical proteins Rv634c, Rv1637c, Rv3677c, Rv2581c from Mycobacterium tuberculosis and to various hydrolases. FASTA scores: sptr|O06154|O06154 HYPOTHETICAL 21.3 KD PROTEIN (200 aa) opt: 355, E(): 4e-15; (37.4% identity in 198 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97145.1"
/db_xref="GI:31619033"
/translation="MAIERVITHGTGTFELDGSGWEVDNNIWLVDGDSVVVVDAAHHA APIIDAVGGRKWAVITGHNDHVTVAPELGTALDAPVLMHPGDVAWRMTHPDKSF RAVSDGDAVVGVTGELRALHTPGHSGVCMWAPELCGTGTGTSFGDTIPAGGPGATG RSVSDFTILRSGLGALPGDVTVHTGHGSDTTIGDEIVHVEEWVARGH"
complement(1998)..3506
/locus_tag="Mb2285c"
complement(1998)..3506
/locus_tag="Mb2285c"
/note="Mb2285c", len: 502 aa. Equivalent to Rv2262c and Rv2261c, len: 360 aa and 140 aa, from Mycobacterium tuberculosis strain H37Rv, (94.7% identity in 357 aa overlap and 100.0% identity in 140 aa overlap). Conserved hypothetical protein, with function unknown but some similarity to N-terminal 70% of P233010|P77703|LNT ECOLI|CUTE|B0657 APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2.3.1.-) from Escherichia coli strain K12 (512 aa). FASTA scores: opt: 239, E(): 1.6e-07, (30.4% identity in 359 aa overlap). Note that neighboring ORF shows similarity to N-terminal part of PCC6803 apolipoprotein N-acyltransferase from Synecocystis sp., suggesting possibility of frameshift. Sequence of clones from two sources has been checked but no error found. Appear to be two extra bases at position 1876970 compared to CDC1551 strain. Conserved hypothetical protein, with function unknown but some similarity to C-terminal end of PCC6803 apolipoprotein N-acyltransferase from Synecocystis sp. Note that next ORF shows similarity to N-terminal part of P74055 APOLIPOPROTEIN N-ACYLTRANSFERASE from Escherichia coli (519 aa). FASTA scores: opt: 142, E(): 0.007, (29.9% identity in 117 aa overlap), suggesting possible frameshift. Sequence of clones from two sources has been checked but no error found.

REMARK-M.bovis-M.tuberculosis: In Mycobacterium tuberculosis strain H37Rv, Rv2262c and Rv2261c exist as 2 genes. In Mycobacterium bovis, a 2 bp deletion (ct-*) results in a single product which is more similar to Rv2262c."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97146.1"
/db_xref="GI:31619034"
/translation="MALRAGARRQPVIGCAALVFGGLPALAFAPAPSWWLAWFGLVPL LLLVVRAPTSWEGALRAWTGGGFWLATQVWLVTSAGPMLVLLAAGLGVWLPAGWL ARLLSVPTTTCVGAALVVVPSVAEAEVWSQSLGPGWALLGASQWFOVTLASA SLGCVLWTSFLVATNTAISVLCVCRATGRLVAGCVIGCAGLSPSYLGSVVRGG PTVRVLVQAGDIADAAARLAAGEFTAAVADQRPDLVWGESVVGQDLTRHPDVLAR LAELSORQADLLVNDAPADPGGIVKSAVLGAHEAVGSKTKRLVPFGVYPLRPL FGMTIRYSKAAAKRQRCQRPVNLAVNSLHIAPLISYEMTSDLTRHAARLGAALLVY QSSTTFGSGWAQPOLAAQAPVRAVEAGIPAVHASLSLSDSSAFDTRGRRLWCSAEFN GAIWVNVPLASNVTLYLRLGDWVPVTFVVMGAGFAVFLRRSLARVSDCAK"
3595..4548
/locus_tag="Mb2286"
/locus_tag="Mb2286"
/note="Mb2286", len: 317 aa. Equivalent to Rv2263, len: 317 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 317 aa overlap). Possible oxidoreductase (EC 1.-.-.-), similar to several oxidoreductases. Similarity suggests alternative GTG start at 10154 but then no zbs. FASTA scores: sptr|O544 05|Q54405 PROBABLY AN NADP-DEPENDENT OXIDOREDUCTASE (297 aa) opt: 487, E(): 1.1e-23; (36.1% identity in 299 aa overlap). Also similar to M. tuberculosis Rv0068, and Rv0439c."
/codon_start=1
/transl_table=11
/product="Possible oxidoreductase"
/protein_id="CAD97147.1"
/db_xref="GI:31619035"
/translation="MAKDLVATVPDLSGKLAIITGANSGLGFLARRLSAAGADVIMA IENRAGAAVEEIRTAVPDAKLTIKALDLSLVAALGQMLADGRPIDLLINNAG VMTPEVTTADGFELOQSGHLGHFALTALLPLLRAAQRAVSVLSLAARRGRHIF FDLQERSYAPMTAYGSKLAVLMEFALDRSRRAAGWIIISNAHPLGLTNTLQIA GSHGRDIPALMERLYKTSWRPAPFQIEIRGILPAAATPOADGGAFTGPRGRY VAVAGGVEAKVPAARNDADSKRLMEVSEQLTGVSYPKSR"
complement(4526)..6304
/locus_tag="Mb2287c"
complement(4526)..6304
/locus_tag="Mb2287c"
/note="Mb2287c", len: 592 aa. Equivalent to Rv2264c, len: 592 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 592 aa overlap). Conserved hypothetical Pro-rich protein, similar to hypothetical proteins Rv0312 (MTCY63.17, 620 aa and Rv0350) that has highly Pro-, Thr-rich C-terminus. Contains P800343 Gram-positive cocci surface proteins 'anchoring' hexapeptide. FASTA scores: Z96800|MTCY63.17 Mycobacterium tuberculosis cosmid (620 aa) opt: 1075, E(): 8.8e-24; (38.9% identity in 627 aa overlap)."
/codon_start=1
/transl_table=11
/product="conserved hypothetical proline rich protein"
/protein_id="CAD97148.1"
/db_xref="GI:31619036"
/translation="MATGAPALGLSIGVTNLAANAADHSITRKPLVLYRPPRPGV LPSVRLDEPGVITDFVDRVDSGVIAGDGSVRSSEALVADALLAYATATGGRA VPSTVTVPAHWGPAVALDALSARASWESHGTSSTAQPLSLPDAALAYAIRAD FGIPARGIVAVCDFGSGGTILVDADRYRPAVTRVQAQFSGDLIDQSLSYVMSE LPQTGAPDAGTSAIGSLTKRTECKAKERSSTVITLTDALGDIKLTNLEEDT IRDLSVSGRALEQTLARGISIRTEALVIVSGGANIPAVITLTLGRCFCVPRTPR PQLTAAGFAGLWAARPPTSATVLTAVTSATAPADAPASVLOPALWSEADESH IGPAGVTAARPSLSFDHAAPEPEKSPIIPWYRLPAVITGTTVALLVGAAGVIG LSTGDOTAGTQRPGVTTTAAAPPSPAPASDGTTEPAPPVQAATGAPPLQOP

LPPTPTNTQPAVTTDVTTPATTPASAPPATTQPPATTQPPATTPSPPPPIPP
IPRIQPPGPIQVPGIGQFSAISGS"
6654..7883
/locus_tag="Mb2288"
6654..7883
/locus_tag="Mb2288"

gene

CDS

Alignment Scores:

Pred. No.: 0.00024 Length: 306550
Score: 666.50 Matches: 629
Percent Similarity: 32.73% Conservative: 291
Best Local Similarity: 22.38% Mismatches: 1094
Query Match: 5.04% Indels: 814
DB: 1 Gaps: 127

US-09-522-753-5 (1-2517) x BX248342 (1-306550)

Qy 20 ProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeu 39
Db 254553 CCCCCGACGGGTTCAGCGC-CCGCGCAGCAGGCGCGG-----ATC 254593
Qy 40 LeuGluTyrGlnHisHisSerArg-----AspTyr 49
Db 254594 CTCGATGTCGGCGCTATAGCCAGCGCTTCTGGGCATCATCTAGGCGGTGCGTGATGCC 254653
Qy 50 AlaSerHisLeuSerProGlySerIle-----IleGlnProGlnArg 64
Db 254654 GGCCACGGCATCTATCAGGCGCGGTGCGGCAGTGCAGCATACACACAGGTT-CGG 254712
Qy 65 ArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluHisLeu 84
Db 254713 CCCCCCGGATGTCATCTGTGTGTCGGCCAGGCGCGGCGGAAACAGCCAGCCCTCTT 254772
Qy 85 -----ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSer 98
Db 254773 GCAGGCGCGCGGCGACGCCAACACGCTGCAGACATGATGCCAACAGCAACCGTGGC 254832
Qy 99 GluMetGluPheIleGluSer-----LysArgPro-----ArgLeuGluLeu 112
Db 254833 CGGTATCGCATTTCTCGCGGTGAACGGAAGCGCGCGTGGTGTGTCATCTCGA-- 254889
Qy 113 LeuProAspProLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySer 132
Db 254890 -----GCTCGCTCTGTGTGACCGCGCTTTCATGTCGCGCCGCCACCGAGGGG 254934
Qy 133 GluAspLeuThrLysAspArgSerLeuThrGly-----Lys 144
Db 254935 CGTCAAGCCGCGCGCAACGCGCGTGTAGTGTGTCGCGTGTGCGCGACGCGCATCGG 254994
Qy 145 LeuGluProValSerProProSerProHisThrAsp-----ProGluLeuGluLeuVal 163
Db 254995 CATCGCGGATTACAGCAGCTTCTCCACGACATCTGCGCGCACCGTCTTAGCAGGCGGT 255054
Qy 164 ProPro-----ArgLeuSerLysGluGluLeuIleGlnAsnMetAspArg 178
Db 255055 AGCGGTTTCGCGCGTGTGTCGTCGTCACACACAGAGAGTTCTCTGCCAGCAGGCTGA 255114
Qy 179 ValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysLeuGln 198
Db 255115 GCAGATCGACACCTCGTAGCGTGAACAGTCACCGCGCGGCGGTGCGCGTTGGCGACCGT 255174
Qy 199 GlnLeuGluGluAlaLysProGluProGluProLysProValSerProProPro 218
Db 255175 CGAGATCAACCCGCTCGGGAAACCGCAGTCCCGA-----ACAGCACCTGTCTCCG 255228
Qy 219 IleGluSerLysHisArgSerLeuValGlnIleIle-TyrAspGluAsnArgLysLysAl 238
Db 255229 GTC-----CGGTGAGCAGCATGTGACCATGTGACCATGTGACCATGTGACCATGTG 255276
Qy 238 aGluAlaAlaHisArgIleLeuGluGlyProGlnValGluLeuProLeuTyrAs 258
Db 255277 GCTGGCGCGCGCACCGCAATACGCGATCCACCGGTGACAGCGGAA---CCGGTCATGCA 255333

Qy 258 nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
Db 255334 AGCTGTGACGATTTTCGTCAG----- 255355
Qy 278 sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPhe 298
Db 255356 -----CGCAGGCGACGACCCCGCAGCG----- 255379
Qy 298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGlu----- 314
Db 255380 -TGC--AAGTTCGATCGCAGCGGATGCGGTGATCGGTGCGAGATCTCGGTCCACCA 255435
Qy 315 -----ArgIleGluAsnAsnProArg----- 321
Db 255436 GGGCGAGGTTGTGCGCAGTGTATCTCGAGTTTCGGCGCGCTCAGCAGCGCGTCCGTGA 255495
Qy 322 -----ArgArgAlaLysGluSerLysValArgGluTyrTy 333
Db 255496 ACAATCTGATCGCTCGCCGTCGCCAGCGGGGAAACCCGCAAAATCTGCTCACCGGCCA 255555
Qy 333 rGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerAr 353
Db 255556 CCGCATCGGTTCCCGCTGCTGCGCAATACCTCAGCGGTGGGACGCCCGCAGCAACG 255615
Qy 353 gValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHis-----GluVa 372
Db 255616 CGACGATCAGAGCGCGCACCGCTCGAGCAAGTCTCGCATGTTCTCCAGCACTACCAACA 255675
Qy 372 lSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLe 392
Db 255676 TGGCGCGGTGCGCGAT-----ACGCGCAAAATGGTGTCCACGTCAGCGGCGCGCGCT 255729
Qy 392 uAlaValIlePro-----ProMetLeuTyrAspAlaAspGlnGlnArgIle----- 407
Db 255730 GATCCGCAACCCCAAAACCGCGCGCGCTCGGCACCGCGCGGTGCGGTGATCG 255789
Qy 408 -----LysPheIle 410
Db 255790 GCGCAGGTTGACATACCAAAACCCCGTCGCGATAACCGTCGCAACCGCGCTCGCACT 255849
Qy 410 eAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAs 430
Db 258850 GTCTCGCAGCGGTGCTTTCGACCCCGCGCACCGGTAAAGGTTGCCACCGCTTTGA 255909
Qy 430 nMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAs 450
Db 255910 CGTC-----CAGCAGCCACCGACTTTCGCCACTTCTGTCACGCGCGCCACCA 255957
Qy 450 nPheGlyLeuIleAlaSerPheLeuGluArgLysThrVal----- 463
Db 255958 G-----CCGAGTCAGCTGGCGCGGAGACAGTGGCAGCAACGACTTTCGGG 256005
Qy 464 -----AlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLy 480
Db 256006 TCCGACGCGCGGAGACCGCTTGTGCAGATCAGGTTGACAGCTGACACCGACCGCTTCCG 256065
Qy 480 sSer-----LeuValArgArgSerTyrArgArgArgGlyLysSerG1 494
Db 256066 GTCGGGCAGGTGCTCAGCGCGGTAGTACCGAGTCTGTTGAGCAGCGCTCTTGGGCA 256125
Qy 494 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
Db 256126 GCAGGTGAGCAACCATGATCTGCTAGTTCGCGCAACACGCTGCTGCGCGCGCGGCGCA 256185
Qy 514 gSerSerGln----- 517
Db 256186 GCTCGCGCACCGCGCGGTGCGGTGCGTTCGATGTCGCGCCCTACGAGTTCCTCGTGGGTG 256245
Qy 518 -GluGluLysAspGluLysGluLysGluLysGluLysGluLysGluLysGluLysPro-- 536
Db 256246 ACAGACCTCCCGGTGTGATGCTGCGATGCGCAGCGGATGCGTGCAGCGCGCGCGCT 256305
Qy 537 -GluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyG1 556


```
Db 256306 GCAAGCCAGGGCGCACCGCGCTC-----GGATGCGGGGCGA 256347
QY 556 uAspAsnAspGluLysGluAlaValaSerLysGlyArgLysThr-----Al 572
Db 256348 ACCGACCAAGAGCTGTCCCTTCGCTTCGAGCGGGCAACCCCGCGGTGCTCGC 256407
QY 572 aAsnSerGlnGlyArgLysGlyArgLysThrArgSerMetAlaAsnGluAlaAsnSe 592
Db 256408 GAACCAATTCGTCAGCGTTC-GGTCAGTTGTCGATCGCGCTGTCTCAAGCTGAGAC 256466
QY 592 r-----GluGluAlaIleThrProGlnGlnSerAl 602
Db 256467 CCGCGAGGTGGGTGCGCGCTCGATATCGCCAGCAGCAAGCTACCGTCCCGTCTCGT 256526
QY 602 aGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGlu----- 618
Db 256527 ACAAGCTCGCTCACCATCTCGCTCCAGTCCACAGGT-ACCACGTTCGAGCGCGGGGTG 256585
QY 619 -----GluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyAr 632
Db 256586 AATCTTGCTCATGTACCAAGCATCGAGCCAGCGCGTAGCGCATTAATCGGCAC----- 256640
QY 632 gAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal-----SerGlnCy 649
Db 256641 -----CTGCGCTAGATTGCTCGAAATCTCTTGGCGCGCGTCCATGTG 256684
QY 649 sLysAsnPheTyPheAsnTrpLysLysArgGlnAsnLeuAspGlu----- 664
Db 256685 TTCTACGCGCTTAGTCATGATGATTCGGCAGCGCGGTGCGATCGCGGCGACCTACAGT 256744
QY 665 ----IleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLy 683
Db 256745 GCGCGTCTCTGCTCGCTCGCAATTGTGAGTCCCCAGTGTGGCAGCATCGCCCGCAGAA 256804
QY 683 sLys-----LysAlaProAlaAlaAsaSerGluGlu 694
Db 256805 CCGACAGCGCATCGTGGCGGTGCGTTCGAGTTCGCGGTGCGCGGTGCGCGAGT 256864
QY 694 la-----AlaPheProProValValGluAspGluGluMetGluAlaSerGlyV 710
Db 256865 CACCGCGCGATGGGTTCGAAAGCTTCCGAAGCTATGATGATGGGGCGCGCTCGG 256924
QY 710 al-----SerGlyAsnGluGluGluMetValGluGluAlaG 722
Db 256925 CCGCTTCGCGTAAACCCCAACCATCGGAAACAGAGCTAGCGTCCCGCGCGCGCA 256984
QY 722 luAlaLeuHisAlaSerGly---AsnGluValProArg-----GlyG 735
Db 256985 GAGCGAGCATCGCGTGGCGCCACACGACGAGGAGCATCGGATTCTCTGACCATGGT 257044
QY 735 luCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProH 755
Db 257045 GGTGTCTCGCTACGTACGTTA-----ACGGAGATGGAGGGCGCGCTTCGCG- 257094
QY 755 isThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProAlaThrLeuGlyA 775
Db 257095 -----CTTCAACCGCG----- 257106
QY 775 laAspGlyProProGlyProProThrProProArgArgThrSerArgAlaProIleG 795
Db 257107 --GAACCGCGGACGAGGTTCGCTCCCTTCGTTGCGGAGCCACCGACTCGCGCGACG 257164
QY 795 luProThrProAlaSerGluAlaThrGlyAlaProThrPro-----ProAlaProp 813
Db 257165 AGCGAGCCCCCGGAGGTCAAGCCACCGCACCGCGTTCGCGCTACCTCCGCGCGCGC 257224
QY 813 roSerProSerAlaProProValValProLysGluGluLysGluGluGluThrAlaA 833
Db 257225 CGTCCCCCGCTACCGCGCGCGATGTGCG-AGGCGGAGGGCGCGCGAGCCCGCGAG 257283
QY 833 laAlaProPro-----ValGluGluGluGluGlnLysProProAlaAlaG 849
```



```
Qy 1171 nLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnG1 1191
Db 258261 GCCGTACCGATGCGCGCGGTACCGCC-:::|||||
Qy 1191 uAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlySerIleThrLysG1 1211
Db 258309 GCGCCC-GCCATCACCGCCAGAGCTTCGGCAGCGGTGCGCGCCCATCACACCGG 258367
Qy 1211 yIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThrHi 1231
Db 258368 CTCGCGCGTCCCCGAATAGCCGCGGTGCGCGGTACCGCCCTGCGCGCGGTGCGCGC 258427
Qy 1231 sGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleLeGlyGluAspSe 1251
Db 258428 CCGCGCGCGCGGC-CTTGGCACCGTTGCC-:::|||||
Qy 1251 rProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrG1 1271
Db 258458 CGACGCGCGCGTCCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 258512
Qy 1271 uGlyLysGlyHisValLeuSerTyrGluGlyMetSerValThrGlnCysSerLy 1291
Db 258513 -GGCGCGGAGGTCCGTTGCCGTGTACCCCCAGACCGCCGCGCGGTGCGCG- 258569
Qy 1291 sGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgThrTy 1311
Db 258570 -----GGSGTCACCGCGGTACCGCGCGGTACCGCGCGGTCCGC 258598
Qy 1311 rAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMe 1331
Db 258599 CGGCGCGCGCGGTGATACCAATGCGCGCGTT--GCGCGCGCGCGCGCGCACCGCGC 258655
Qy 1331 tGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisI1 1351
Db 258656 CCGCGCGCGCGA-GTCCGCGGTGCCACCGACCCACCGCGCGGTGCGGTGCGGTGCGCGA 258709
Qy 1351 eArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLe 1371
Db 258710 -----CGTCCCGAAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258744
Qy 1371 uArgArgGluAlaLysLeuLysArgGluGlyThrProProProProProSerAr 1391
Db 258745 TCCCGCGCGCATGTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258803
Qy 1391 gAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLysLeuLysProAlaHi 1411
Db 258804 ACCGTGCGCGCGATCTGTGATA-CCGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258843
Qy 1411 sGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgG1 1431
Db 258844 -----GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258854
Qy 1431 uGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly---Se 1450
Db 258855 ACCGTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258910
Qy 1450 rIleThrGlnGlyThrProLysTyrAspThrGlyAlaSer---ThrThrGlySerLy 1469
Db 258911 CGTACCCGACACCGACCC-----ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258956
Qy 1469 sLysHisAspValArgSerLeuIleGly-SerProGlyArgThrPheProProValHisP 1489
Db 258957 -----GCGCGCGTCCCGCGCGTCCCGCGCGTCCCGCGCGTCCCGCGCGTCCCGCGCG 258994
Qy 1489 roLeuAspValMetAlaAspAlaArgAlaLeuGluAlaCysTyrGluSerLeuL 1509
Db 258995 CC----- 258996
Qy 1509 ysSerArgProGlyThrAlaSerSerSerGlySerIleAlaArgGlyAlaProValI 1529
Db 258997 -----GTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259020

1529 leValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaP 1549
259021 -----GAACGAGGCGCGCGCGAAATTGGCCCGTTTGGCGCGCGCGCGCATCGCCAC 259072
Qy 1549 roPheAlaGlyHisLeuProArgGlySerProValThrMetArg-----GluProT 1566
259073 CA-----TTCCCGCGCGCGCGCACCATCGCGACCCCGCGCGCGCATGTCGCGCGT 259120
Qy 1566 hrProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAspArgLysLeuT 1586
259121 CGCTCCGTTGCC-----C 259135
Qy 1586 hrSerThrProArgGluLeuAlaLysSerProHisSerThrValProGluHisHisProH 1606
259136 CCGCGCGCGCGCGCGCGTTCGCGACCGTTCGCGACCGTTCACCGTTCGCGCGCGT 259195
Qy 1606 isProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgS 1626
259196 TGGCG-----GGCCCGGTGG-----TCGCTGGCGGTCCACACCGCTG---GCCG 259234
Qy 1626 exHisIleProLeuAlaPheAspProThrSerIlePro-ArgGlyIleProLeuAspAla 1645
259235 CGCGGTGCGGTTCGCGCTGTCGCGCGCGCAAG---CGCGGAGGCGCTCCCGCGCGCGGT 259291
Qy 1646 AlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr 1665
259292 CACGCGCGTCACGCGCGTCCCGCGCATCA-CTGCCATGACCGCGGTCCACCGCTTGCGC 259350
Qy 1666 Pro-----ProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn 1681
259351 CCGGTCTCCGCGCGCGCGC-GTCCACGCCAACCGCGCGCGCGCGCGCGCGTGGCT--- 259400
Qy 1682 ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisAsnThrAlaThr 1701
259400 ----- 259400
Qy 1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla 1721
259401 -----GCCGACCTGACCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCATC--- 259442
Qy 1722 LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro 1741
259443 -----ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259454
Qy 1742 ValLeuVal-ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLe 1761
259455 -----GCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259491
Qy 1761 uProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyG1 1781
259492 -CCTTTCCCGCGCGCGCGCTCCACGCGCTCAGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCG 259545
Qy 1781 yProThrHisLeuThrLysProThrThrThrThrThrThrThrThrThrThrThrThrVa 1801
259546 -----TTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259574
Qy 1801 pArgGluArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrVa 1821
259575 -----ACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259621
Qy 1821 lGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSerGlySerSe 1841
259622 CCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259654
Qy 1841 rGlyGlyGlyGlyGlySerSerSerArgProAlaSerHis----- 1854
259655 C-----GGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259707
Qy 1855 -----SerHisAlaHisGlnHisSerProIleSerProArg----- 1866
259708 CCACTACCTAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259767
Qy 1867 -ThrGlnAspAlaLeuGlnGlnArg---ProSerValLeuHisAsnThrGlyMetLysG1 1885
```


DEFINITION Streptomyces verticillus bleomycin biosynthetic gene cluster, complete cds.

ACCESSION AF210249

VERSION AF210249.1

KEYWORDS GI:9937210

SOURCE Streptomyces verticillus

ORGANISM Streptomyces verticillus

REFERENCE 1 (bases 1 to 74593)

AUTHORS Du, L., and Shen, B.

TITLE Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer Streptomyces verticillus ATCC 15003

JOURNAL Chem. Biol. 6 (8), 507-517 (1999)

MEDLINE 99352421

PUBMED 10421758

REFERENCE 2 (bases 1 to 77457)

AUTHORS Du, L., Chen, M., Sanchez, C. and Shen, B.

TITLE An oxidation domain in the BimII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in Streptomyces verticillus ATCC15003

JOURNAL FEMS Microbiol. Lett. 189 (2), 171-175 (2000)

MEDLINE 20389599

PUBMED 10930733

REFERENCE 3 (bases 1 to 77457)

AUTHORS Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.

TITLE The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase

JOURNAL Chem. Biol. 7 (8), 623-642 (2000)

MEDLINE 20500448

PUBMED 11048953

REFERENCE 4 (bases 1 to 77457)

AUTHORS Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.

TITLE Direct Submission

JOURNAL Submitted (01-DEC-1999) Chemistry Department, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA

COMMENT On Aug 29, 2000 this sequence version replaced gi:5326869.

FEATURES

source

1. .77457

/organism="Streptomyces verticillus"

/mol_type="genomic DNA"

/strain="ATCC15003"

/db_xref="ATCC:15003"

/db_xref="taxon:29309"

complement (1199..2185)

/codon_start=1

/transl_table=11

/product="SyrP-like protein"

/protein_id="AAG02342.1"

/db_xref="GI:9937211"

/translation="MTGSVLTPLPGIIPRPROGELTGAEYDLGDLGADGPDWRAH GPRLRLATDGLIHLLGPTDGDGVGDDVGVGGDPLPYTERSTRVSVKGIY RSTEYPADOP IPMHNENSYAAHWSPSLYFFCHTAPDTGATPIADGRAVLDLPABEV RFSQGVVYTRFRADWGLSMQAFOTEDRGVHCRAGQFSDGVDLVRHRHP RRAVDPTGAEVWFNOAHLPHPSLDPDLRQVLLETYGENGLPRDLAFADGTIPDAD LATVRAAYTAALALPWREGDIMLVNLRNMAHGREPFTGERRVLVMTSADS"

complement (2182..3918)

/codon_start=1

/transl_table=11

/product="peptide synthetase NRPS12"

/protein_id="AAG02343.1"

/db_xref="GI:9937212"

/translation="MTLTLDADFDAQPTPDHAHVHGTDTVMYRELELRACRMART LAARGAGPTLVAVRPGPEPVAALLAVLITGAGVPLADDPPDCRHLDDCAA LLIHAHPSRDGRLTTPDEALAPAPFDAAPVRAGDPAYVITYTSSGRPKGVLEQGA LGAYLAQARAYDGLSRTVLHSSLSLSDMAVTSIMGLVSGGAIHVLDLKAIASGTQP PPAASRPSFKTPTPSHLPLGLIPDSCLPTGQLVIGGENTLGSALGPWRAHPDVTY VNSYGPTEATVGCATVRPGDAVDAPGPIGFPFAGTRLDYLDADGPEVAVGVGVEL HIAGDQALGKYLRLTEERFPDPFADGSRMYRTGDLVRERPDGLEYLGRADGQ VKVSGYRIEPPGEIEAVLRGHAGVRDCAVAVGADARRLVAYVVPDPSPPGTAAPAR

CDS

complement (3915..4733)

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG02344.1"

/db_xref="GI:9937213"

/translation="MSKNRAHWSRITGDAPGVVLAVDFYGTGRQEATFPHLCDLLTD PEVHAPVPAPDGDWSTAGHLRWTEGLTVLAGRPVRLAGVFCAGGVFASAIRA DALVREGRHPRVLPFNAPGATVTRDFRGLIAGMDLLDGERAALLAETTAIRRA HAPDALVPAERYAALYREGCDLLCERLGVDSFGAELAAVLSYLAYLTAALDVPPT PLWRGAVSLTSTREHQGTFTDVEHGFDAVARELLSSQVVAALTALLREHASR"

complement (4741..6024)

/codon_start=1

/transl_table=11

/product="putative transporter"

/protein_id="AAG02345.1"

/db_xref="GI:9937214"

/translation="MTARVDATPTYLAVLAVREARAPLLGSCILARMSFVLPPLALLIS VRDATSPVAGLTSGALSNTLTFAPARARLIDRGSRGLVRLTPYLLGLAVLIT LAEEAPTALLVAAAVAGVAPPGLPTMRVLWARILHGRQPLHTAIGADESVTEEV FTVPGLLAGGLTAAAPLASMITVMVLIAGTACFVLSAATAAPASGAEDEDRPHGR PMALPGMRTIVISFGVGLVGVLPVLPFIADHAGSPGAGGILLMSLGSAGVGGLA YGRIAWRSTPVRFRVVLVTGLVPLCLTASPVAGAPALLVGLCLAPLFTTAYLL VNDLVTASGTAPTEANTWSTANNGGFAAGSAAAGVLLDSRGFTVTVTAAFAVAAAAT VMTVLRRTLLGAGHPPEAATPADRTAPAEAB"

complement (6021..6773)

/codon_start=1

/transl_table=11

/product="thioesterase"

/protein_id="AAG02346.1"

/db_xref="GI:9937215"

/translation="MDPFLTRVNPWFSGGCDGRPRVRLCALPYAGGTAAVFKWPAAL PGVELLTAHLPGRGDFTTEPPATLEAERLCEALPPSDLPVLVHGMGALLGYE VAARLAARGAPNLIIAACRPPHPDASGPVTEAEALATRAERPWDLTALRDEELM EALVPALVADI TAGDRYHRPRRPLDLPLKVIYICADDGDTWRTTLGWRACRTADCEV VVLPGGHYFLETDRAAVLRATVADLAEEVGA"

complement (6784..7797)

/codon_start=1

/transl_table=11

/product="clavaminate acid synthase-like protein"

/protein_id="AAG02347.1"

/db_xref="GI:9937216"

/translation="MTDLRLRTVALTGESEAVDDLLRLTLADVPDSTVGLLHRLTALA QELPRLRAELTGRLYDPSRALVVTGVDDEIRIGTTPAARPAAPDPPTRLDELLL LHAALLGEAFGWATQOQRLVHDVLPVGEETAQMGSSSETELLWHTEDAFHLRCD YVGLLCLRNHQRAATTVGWPDLSRLTTEDEVALLPEPRYLIRPDTSHTPAQATGTRSA ERFAAIAEMDDAPERVALFGDPEDPYLRIDPAYMSPAGDAAARAYDTVTALIDE LRVHVLDAAGSLLLVDNYQAVHGKPFAPAAAYDGRDRLKRVNITDLRRSSRRSATS LLV"

complement (7859..9214)

/codon_start=1

/transl_table=11

/product="putative transporter"

/protein_id="AAG02348.1"

/db_xref="GI:9937217"

/translation="MMKSRLRDRQLGGEDPVVAQSPQDAGTTCQGGDGLNVFAAL AALLEVEVPVRLPHHAGLGRHVHSSQAAVALGPMQVASATTGVAGDGHQAGRRQO VTGCVGVAGVAGDDELGAEDRPHARQRLDGLGMAERLADLLNPLQTVVQGOOL RGQVSHDLGCDVLPQGRGLSLGSLQRRRGGVGTAAHSVQPGRQPGPAAAECSRG LVARQODQALLRAVVEGPFQGEDAGQVATVDHPDPVGHVGTGGQGLVAGQGOOL GHVDRRREPVAGGDDVGVAGVGLRLAPVSAHVAADRAAGHVDGLLAVGREGQQO QKGRAGDVHCPVDLVGVQVEDLTHGGQRRLLIVADLLRPOQRHTLVLDHRRPVMPLARV QEPGDPGSLAHSPLLTPTSIPSTRGTPTCHLKKRPKRTSQSAAARAPRTRGRPLVSH"

complement (9211..17736)

/codon_start=1

/transl_table=11

/product="peptide synthetase NRPS11-10"

/protein_id="AAG02349.1"

```
/db_xref="GI:9917218"
/translation="MASDLSAHLVSDCRORGGITVTNDGSELAQNVAAVRPRYS
ALAPERTALHKGAGTGYDELNRRAELTATRLADAGAGSTLVAVALPRDPVLATLC
ALLKGAACLPIDPGIPAGRLREIMADASPLDVTTRVAPAFPGDGPVFLDDAPPT
CAVILPRHSAGTSEIAYVLYPTTDEKSENSVSDRYMDLDDPTAGIPARABILR
LVAVLPSGLVLDADDETPRVPTREAPRDMVDVVAQVCAVLGVVGRDRFDL
GKSLAAOVAVARLKLGLVELPLRALFDPTVEELAAVRAEQAGGQGVREAALEP
VGRSEPLPSAQORLWFLDRLMPORAFYTMCDAPRVGGIDGLRALRALMLVGRHE
TURTAVERDGPYQLVGPADGARGARVAAPTRVLSLLEPAEREAVRNLVAAART
PRPADGALLRVVVARLADDDHLVTVSTHIIYKSDAWSGVVLDELGLRECVTGDPA
ALPPAVQAVLAWQORAWMGVQOEHLAYKRALDGSPVLRILFMDHPRPAVOSR
GETVGFALPDALVAALCKLGRQCATLFTLLAQFVLLARHAGQEDIVVGVPAAGT
RTETEPFLVDLPLRLAI CAQGLSFDRLLDQVREAAALCAFAHODLPEALVEALAP
ERDLGNPLVQVTFQLLCTPAARPDLICTEVERYVQOEAVSOFDLSLDKRADDSGR
GIUNYCPDLFRRRMEVVLGHYTLTLGAAADPGRI GELPLSDGNERLRLLDGPKR
DRAYAGFVSPERFAEVARTADARATTCATTLTFAELNDRVERLALQALDGVRE
TPVAVLRPSTDSVALLAVMRAGGVVPLDPMDPADRTAYILDDTASVVTIRDLPA
LPGLRHVDRPADGVPAPRIDPDQAAVVIYTSGTGAPKGVVVRHRSLSLHLSAL
QATFLGHPVLACAGVPPGDAKLRITLTAPFTFDASMBQLSMWLAGHELFIVPEDVR
RDSALRVFREHRIIDVITTSQELLSHGLLDGEWAPSMVMVGGEAVSPSLMRTL
RDQRTCFNLGYPTATVDATCDLSDPADVPVIGTPLPHTRVRLDRLRPVGVG
AGSIYLGTLGRFVRLPALTARRFVADPYPDTGSRILYRTGDRMRPDTGLEITG
RTDQIKIRGFVPEGEIEAVLTHPFAKEAAVDDAHARLVAYVTLAEGGAGPTDV
RRPAQGLPAHVPVSVVLEALPLTSGKLLDRALPAPAAGRPELDFVFPADRMVE
EVVAQWCAVLGVDRGVGDDHDPFELGSHSLVVQVMTIRKLLGVPEVPLRELFDAATV
EELAAVRAARTEGLRGAAPPLGVPDRSGPLSLFAQORLWLDOLA PDSVYNMCD
IYAGRTDLDLURKALRILVERHETLRTAFVERDGVPHQVVSAPPAPARAEEVVR
AEAGRTDEVRDLVAABETPRPADGALMRVAVARLADDDHLVTVSTHIIYSDGNS
VDLILDELGLRYHRTGDPAGLPDQVAVFVWQSWMTGPFVREHLAYWKRALD
GAPSLRLPADHPRPAVQSQRGETVEFFLPAPLVARLEALCRQGVTLFMAFLGFQV
LLARYSQDDVPLAGVPTANTRAETPLVGFVNTPLVRVACSPSLFRALLDRVREA
ALGAFHQDLPLFAALAPALAPEDLGHPLVOVTFOLLADDERLVLHGTDCVSLGFG
GVTSFPLSDVSVSGRGRKCVITYCPDLFDRPRMEVLGHYTLTLAGAADDPGLRVG
DLPLSDDDVERLLGSGRPYLPAGAEIVPDPAQVRAVTPADPALVHGDSTLTFAE
LDTRVTAALVRLRCRQVAAETPVACLPRSDADVALLAVLRAGGVYVFPVDEPWSGR
VAHVLDTAAPVYITRLDPADPGRVHLDPRQAPADDDPLRLHRDQAAVYIIFTSST
GAPKGVVHSGSLVHLGHVRRMAEGPRNVAHTTAMTFDPSLEQFLWLVAGTILHV
APEVEDRPAALVALPRAAI DVLNVTTPSHLTLI EAGLEGGDRVGTVLVGEAVPA
ALMRTLHRTGRTAFNLTGPTATVDATCDLSDPADVPVIGTPLPHTRVRLDRLRP
RPVFGVAGSEIYLGTLGRLYLNRPALTAQRFVADPYPDTGSRILYRTGDRMRPDP
GTLEYLGRTDQIKIRGFVPEGEIEAVLTHPFAKEAAVTDGAKRLVAVPVA
PRAPHGDSADGADPAQVEEWNVAFEATHTDAADGELTFNKGWNSLTGAPLPAEHMR
EWDVTVARLLERPAERVLIEGSGTGLLWRLLPHVTEYTGTFDSRPVMDLRDGLR
RPAHRLVLRHREATDFGVRAASTDLVWVSVQYTPDRAYLDTLVARALDADADRGR
VFGDVRNLALAQFYAROLAHAGPCGAARDVARAAGEFAAMDGELLVSPAFAFLA
ARSPRVTVETILFRGRHNMESLYRDVVLVHVGDRPAPAEAVLTWGDQVHDLSL
SARLGRGPDALLVRGANDRLTRDNELLDAPARTTAVEPEDLWGLADSTPQVSVSW
AAADPRGAMDVLLVRRDADDDGPLLVPHPPEPSAPLTNTPTTRHPSARQGSAAAGLR
SWLAERUPALLPARI TEVDALPRTGTGLDRGALGLVTAGRGARAGDRPATAPRTG
LERTLADAMARVLGPEVGVHENFVALGGSLLAVRAVARCRRAGVRLTVRQLLSBQT
VAALAAALEESQ"
complement (18307, .18660)
/codon_start=1
```

Alignment Scores:

Pred. No.:	8,11e-05	Length:	77457
Score:	566.00	Matches:	681
Percent Similarity:	30.64%	Conservative:	275
Best Local Similarity:	21.83%	Mismatches:	1098
Query Match:	5.04%	Indels:	1078
DB:	1	Gaps:	156

US-09-522-753-5 (1-2517) x AF210249 (1-77457)

Qy	3	GlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrProProHis	22
Db	64608	GGTGGCGACGCTCCCGCTGCTTCCAGCTACGACCA-----CCGCTCGGA	64661
Qy	23	SerLeuSerTyrProVal-----GlnIleAlaArgThrHisThrAspValGlyLeuLeu	40
Db	64662	CGAGTCGCGACCGGCTGACCTGAGCTTGGACCTGCTCAT-----	64703
Qy	41	GlutyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleGln	60

Db	64704	---CGCCGACGCCCCACAGCGTCCACGTACTACCGCGGACCTGCTCACCCTTCTACGC	64757
Qy	61	ProGlnArgArgProSer-----LeuLeuSerGluPheGlnProGlyAsn	76
Db	64758	CGACCCCGCGCGCCCTGCGCCCTCGGTGCTCTTCTCGCGACTAGCTCTGCGCGT	64817
Qy	77	GluArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGly	96
Db	64818	CGCGC-----CCAGCCGAGGG	64835
Qy	97	LysSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuLeuProAspPro	116
Db	64836	CGA-----GCCGCGCGCGCGC-----CCTCGACCA	64862
Qy	117	LeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThr	136
Db	64863	CTGGCGGCGCGGCT-----GCCACCTGCCCG-----	64892
Qy	137	LysAspArgSerLeuThrGlyLysLeuGluProValSerProProSer-----	152
Db	64893	-----CCGCGCGCGCTGCGCTGCGGTGCGGCG	64922
Qy	153	-----ProProHisThrAspProGluLeuGluLeuVal	163
Db	64923	CGAGGAGCTGACCGCGCGCGTTCGCGCGCTCACCCCGGACTCGG-----	64970
Qy	164	ProProArgLeuSerLysGluGluLeuIleGlnAsnMetAspArgValAspArgGluLeu	183
Db	64971	CCCCACGCGCTGGGCACGCGCTCG-----GCCGCGCGCGCGCGCGCGGCGGCGG	65018
Qy	184	ThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnGluGluGlu	203
Db	65019	CACCCCGCGCGACTGATCTGCGCGCTTCTGCGAGCTCTCGCCAGCTGAGCGACAC	65078
Qy	204	AlaAlaLysProGluProGlu-----LysProValSerProProIleGluSer	221
Db	65079	CCCCCGCTTTCACCTCCTCACCTCACCCACCTTCCACCGCGCGCGCTGCTCCCGCGGTGA	65138
Qy	222	LysHisArgSerLeuValGlnIleLeuTyrAspGluAsnArgLysLysAlaGluAlaAla	241
Db	65139	CGACTCTGCGGACTTAC-----CACACGACCTGCTCGGGTCGACGCGGAGGGGA	65195
Qy	242	HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer	261
Db	65196	CAC-----CTTCGCGGACCGGCGCGCGCTCCAGGACCGCATCTGGGAGGACCTCGA	65249
Qy	262	AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle	281
Db	65250	ACACCGCGCTGTCAGCGCGCTGAGGTCTCGGATGCTGCG-----	65291
Qy	282	LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGlnArg	301
Db	65292	-----CCGCGACGCGGCGCACCCACGACG-----CGT	65318
Qy	302	TyrAsp-----GlnLeuMetGluAlaLeuGluLysValGluArgIleGluAsnAsn	319
Db	65319	CCGGATGCGGTGCTCTTACACGACCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG	65378
Qy	320	ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu	339
Db	65379	CCCGCGCGCTGCGCGGTACGCGCGCGGTACGCGGTACGCGACCGACCGCGAGTCTCT	65435
Qy	340	IleArgLysGlnArgGluLeuGlnGluArgMetGlnSerArgValGlyGlnArgGlySer	359
Db	65436	-----GCTCGACCATCAGGTGAGCGAGAGCGACGCGCGGACT	65471
Qy	360	GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu	379
Db	65472	GGTCTGACCTGAGGACTACGTCGCGGACG-----	65501
Qy	380	SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProMetLeu	399

Db 65502 -----CTACCGCGCCCG--- 65513
Qy 400 TyrAspAlaSerGlnArgIleLeuPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Db 65514 -----GCTGATCAGGCCATGTCGG-----GGCTTCGAGGCGCTCCT 65552
Qy 420 MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr 439
Db 65553 CGCTCGCTCGCGGTGTCACGACGA----- 65576
Qy 440 PheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGlu 459
Db 65576 ----- 65576
Qy 460 ArgLysThrValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyr 479
Db 65577 -----CGACCGCGCCACGACGA 65594
Qy 480 LysSerLeuValArgArgSerTyrArgArgGlyLysSerGlnGlnGlnGln 499
Db 65595 CGACCGCGGCACGACGCGCCCGCCACGACGAGCGCGCCCGCCACGACGCGGCC 65654
Qy 500 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
Db 65655 CGGCCACGACGCGCGCCCGCCACGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 65714
Qy 515 -----SerSerGlnGluLysAspGluLysGluLysGluLysGluLysGluLysGluLys 530
Db 65715 CGGCCACGTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 65773
Qy 531 LysGluGluLysPro-----Glu 537
Db 65774 CGGAGTCAGACGCGCGCGCCCGCCACGACGACGACGACGACGACGACGACGACG 65833
Qy 538 ValGluAsnAspLysGluAspLeuLysGluLysThr----- 550
Db 65834 GCGGATGATGAACGACGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 65893
Qy 551 -----AspAspThrSerGlyGluAspAsnAspGluLys 561
Db 65894 CGACCGCGCGCGCACGACGACGCGCGCTCTGGCACTGGTGACCGCGCTGACACTGAG 65953
Qy 562 Glu-----AlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArg 577
Db 65954 CCACGCGAACTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66013
Qy 578 -----Arg 578
Db 66014 CATCCCCGGGACCGGGTCGCCCTCGCGCTCGAGTACGGCTGGGAGCAGGTCTGGGCGC 66073
Qy 579 LysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluAlaIle-ThrPr 598
Db 66074 CTGGCCGCGCTCGCGCGGAGCGGTCTGCTCGCTCGCGCGCGCGCGCGCGCGCGCG 66133
Qy 598 oGlnGln-----SerAlaGluLeuAlaSerMetGluLe 609
Db 66134 CGCCCGCTGGCAGCACGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66193
Qy 609 uAsnGluSerSerArgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuG 629
Db 66194 CACCCAGCGCATCGACTGGCGCGCAGGAA-----CTGCCCGCTCTCTC 66235
Qy 629 uHisGlyArgAsnTrpSerAla-----IleAlaArgMetValGlySerLysThrValse 647
Db 66236 -CGTGGACGAACCG 66294
Qy 647 rGlnCysLysAsnPheTyrPheAsnTyrLysArgGlnAsnLeuAspGluIleLeuG 667
Db 66295 CGACCGACGCGCGCTACCGGTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66343
Qy 667 nGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLysLysAlaPr 687
Db 66344 -CACCAACCGCGCGCTCGGAGATCGACCGCGCGCTTTCGCGTGGACCGCGCGCGCG 66402

Qy 687 oAlaAlaAlaSerGluGluAlaAlaPheProValValGluAspGluGluMetGluAl 707
Db 66403 TGGC-----CCTGGCCC-CCGCCGACTCGCGCTCTCTCTACGAACTG 66446
Qy 707 aSerGlyValSer-----GlyAsnGluGluGluMetVa 718
Db 66447 TTCGGGCCCCCTCTCTGGCGGTGGCGCTCTCTCTACCGCGGACATCGACTCGCGAT 66506
Qy 718 lGluGlu-----AlaGluAlaLeuHisAlaSerGlyAsnGluValPr 732
Db 66507 CCGGAGCCCTGACGAGGCGCTGCGCACCCACGCGCTCACCTCTGGCACTCGCGCCC 66566
Qy 732 oArgGlyGluCysSerGlyProAlaThr----- 741
Db 66567 GCCCTCTCGCGCT 66626
Qy 742 -----ValAsnAsnSerSerAspThrGluSerLleProSerPr 754
Db 66627 CTCGGCTGTGTCTCTCGCGCGGAAACGCTCTGACCCCGCTCTCTCTCTCTCTCTCT 66686
Qy 754 oHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuG 774
Db 66687 GAGAGCCCCCGCACGACGCGCGCTGCGCACCTCTCTCTCTCTCTCTCTCTCTCTCT 66738
Qy 774 yAlaAspGlyProProProPro-----ThrProProArgArgTh 789
Db 66739 -----CCGGCCCCCTGACCACTGCTGTGAGACCGCGGACCTCGCCCCGGAATGG 66788
Qy 789 rSerArgAlaPro---IleGluProThrProAlaSerGluAlaThrGlyAlaProThrPr 808
Db 66789 CGCTCGTCCCGTTCGCGCGCGCTCTGCGCGCTGTGCGCGGCGCTGATCGACGTCTC 66848
Qy 808 o-----ProProAl 811
Db 66849 CTGCGGCGCTCGCGGTCTGGGTACCGCGCGCTCTCACTACGCGGCGGTCTCGCGCGAG 66908
Qy 811 aProProSerProSerAlaProProValProLysGluGluLysGluGluGluT 831
Db 66909 CCCCCACCGGAGGAGCAGCACCG---CGACCGTCCGCGACCGCGGAGCGCGGAA 66965
Qy 831 hrAlaAlaAla-----ProProValGluGluGly----- 840
Db 66966 CCGTGTGTGCGCACCGCGCTGTTCGCGCGCTGTGCGCGGCGCTGATCGACGTCTC 67025
Qy 841 -----GluGluGlnLysProProAlaAlaGluGlu----- 850
Db 67026 GCGACGAGACCGCGCGATCAGCGTCCGCGACCGCGCTCTGAACTTCCAGGACACCGAG 67085
Qy 851 -----LeuAlaValAspThrGlyLys----- 857
Db 67086 ACCGCCCTCGCGCCACGAGGAGGTGACATTCGCCCGTGTCTCTCTCTCTCTCTCTCT 67145
Qy 857 ----- 857
Db 67146 GACGAGTCTGCGCGGTGACGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67205
Qy 858 -----AlaGluGluProValLysSerGluC 866
Db 67206 CTGCCCCATCTGCGCGCAAGGTCTCCCTTACTCTGCTCGCGCGCGCGCGCGCGCG 67265
Qy 866 yThrGluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 885
Db 67266 GTTCGCTGCGCTCACCCCGGACCGCGCGGTGGACCGCGCGCGCTCACCGCGGAGGC 67325
Qy 886 ThrAlaGluGlyAlaLeuLysAlaGluLysGlyGly-----SerGly 901
Db 67326 CCGCCCCC---GCTGC-CGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67381
Qy 902 ArgAlaThrThrAlaLysSerSerGlyAlaProGlnAspSerAspSerAlaThrCys 921
Db 67382 CGAGCGCGAACTCTCTCGCCCAAGTGGC---CCGGGTGACTCTGCGC----- 67423

QY	922	SerAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArg-LeuLeuSerProAr	941	Db	68397	GACTTCA	: : :
Db	67424	-----GGTGTGGGAATCGCGCGTCCGAACCCGA	67453	QY	1218	-----ProSerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAs	1236
QY	941	gProSerLeuLeu-ThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProL	961	Db	68442	CGGACCGCGTCCGACCGGATCGCCCTGTCGCCCTCGCACAGATGGACGCCCCGCGCG	68501
Db	67454	TATGAACCTCTCGAGCGCGGTG---CCACCTTCGTCGAACACTGTCGCGCTCGGACCGC	67510	QY	1236	pValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAs	1256
QY	961	euAspLeuLysGlnLeuLys-----GlnArgAlaAlaIleProProI	976	Db	68502	CTGCGCCGCGGTTCGCCCTCACCGACCGGACCGCTTCGTCACGCCCTCTCTCGGCGG	68561
Db	67511	TCG-GAGGAGAACTCGGCTCGACACGACATCGAGGAACGTGTGGCTTCCCGTCGG	67569	QY	1256	p-----ArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGlu--	1271
QY	976	legInValThr-----LysValHisGluProProArgGluAspAlaA	990	Db	68562	CGGCCCCGACCGGAGGCCCGTGAACCGCACCGGCCCTCGCGCGCGCGGAGACG	68621
Db	67570	TCGCGGTATCGTCGCGCGCCACTCGCGCGCGGACCGCACCGCGCGCGGACCCCG	67629	QY	1272	-----GlyLysGlyHisValLeuSerTyrGluGlyG1	1283
QY	990	lapro-----ThrLysProAlaProProAlaA	998	Db	68622	TCGACACCGACCGCCGCGTGGTGGCGCGCACCGGCGACCCCGGGGTGCGGTGGAGG	68681
Db	67630	TGCGCGCGCGTTCGTAGCGTTCGACCCGCGTCCGTACTGCGCGCGCGCGCGCGCG	67689	QY	1283	yMetSerValThrGlnCysSerLysGluAspGlyArgSerSerGly-	1299
QY	999	-----ProProProGlnAsnLeuGlnProGluSerAsp-	1010	Db	68682	T-----GGCGCGCACCGGAGCGGCGTGCACCGCGCGGAATGGCGCC	68723
Db	67690	GACCGTGGCGCGCGTTCGTGCGCGCGCGCACCGCGTCCGTACCGCGCGCGCGCG	67749	QY	1300	-----ProPro-----HisGluThrAlaAlaProLysArgThrTyrAspMetMetG1	1315
QY	1011	-----AlaProGlnProGlySerSerProArgGlyLysSerArgProA	1027	Db	68724	GCACCCACCTCGACACCGTCAACCGCTGCTGCACCGTCAACCGAGC-	68769
Db	67750	CCTCACCGCTCGCGCGCGCGCACCGCGCGCGTGCACCA-----CCACCGCGCG	67797	QY	1315	uGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaI1	1335
QY	1027	laproProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuLysLeuProGlyAspProp	1047	Db	68770	-----CGTCTGT	68777
Db	67798	TCCGCGCGCGCTCGTCCCGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG	67857	QY	1335	eProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGlySerI1	1355
QY	1047	roCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLysAlaSerP	1067	Db	68778	TCGCGCGTTCGCGTTCGCGCTCGACCGCTTCGCGAGCTGCTCCACGCGCTTCGCGGAT	68837
Db	67858	CGCA-----CCCATCCCGCGCGCTC-CGT-----GCC	67886	QY	1355	eThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArgGluAl	1375
QY	1067	roHisAlaProAspProSerAlaPheSerTyr-----	1077	Db	68838	CCCCGAGCGTACGTGGAACCGTCTGCCCGCGCAC-----CGCCCTCGGCG	68895
Db	67887	CCCG	67946	QY	1375	alysLeuLeuLysArgGlu-----GlyThrProProProProProProSerArgAs	1392
QY	1078	-AlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuP	1097	Db	68896	ATCACCCTTACACCGCGCACCGACCCCGCGCGCACCGCCATCCCGCGCACACGAGA	68945
Db	67947	GACGCCACACCGCATCCGACGAGTTCAGCGCCACCGAGCGGTGCGCTCAGCGGC	68006	QY	1392	pLeu-----ThrGluAlaTyrLysThrGlnAlaLeu-----	1402
QY	1097	ro-----ArgProPro-----ThrIleSerAnProProProLeuIleSerSerAlaLysH	1114	Db	68946	ACTCTTACCAACTCCGCTTCCCGGAGCGTGTCTTTCGGCTGCCTCACCCCGCGCGGA	69005
Db	68007	CGGACGACACCACTCACCGCGCGCTCGAGCGCACCGACCGCTTCGACCCCGCGCGCGT	68066	QY	1403	-----GlyPro-----	1404
QY	1114	isProSer-ValLeuGluArgGlnIleGlyAlaIle-----	1125	Db	69006	CGGCGCGCGGACCCCGCTCGCGCACACCGCGCGGTCTCTGGCGCGCTCGACCCCGCGCC	69065
Db	68067	ACGTGCGCGACCTGGCGCGCTCTCTCGCGCGCTCTCGCGCGCTCGCGCGCGCGCGCG	68126	QY	1405	-----LeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyAr	1423
QY	1126	-----SerGlnGlyMetSerValGlnLeuHisValProTyrSer	1138	Db	69066	TCGTGCGCGCTTCGCGCGCGCGCGGTCTCTTACCGAGCAACTACGCGACGGATCG	69125
Db	68127	GAAACCAATACGCTATCTCGTGGCGGTCTCTCTC-----TACCGCTCCAG	68174	QY	1423	gSerIleHisGluIle-----ProArgGluGluLeuArgHisThrPr	1437
QY	1139	GluHisAlaLysAlaProValGlyProValThrMetGlyLeuPro-----	1153	Db	69126	G-----CATGCTCTGGCAGCGCTTCCAGACCGCGCACAGAGCGCGCGCTCACCG-CC	69178
Db	68175	ACCTACCTGCTCTCCACCGCGGGAAGGTGAC-----GAGCTGCCCGCGCGCGCGCG	68231	QY	1437	oGluLeuProLeuAlaProArgProLeuLysGluGlySerIleThrGlnGly-----	1454
QY	1154	-----LeuProMetAspProLysLysLeuAlaProPheSer	1165	Db	69179	TACTGCGCGCGCGCGC-----CGGTGACGTGATGGAACCCCGCGCGCGCGCGCGGACC	69235
Db	68232	GTCCACCCCGCGCGCAACCGCTGTGTAGCATCGACCCCGCGCGCGCGCGCGCGCG	68291	QY	1455	-----ThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysHis	1471
QY	1166	Gly-VallYsGlnGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuG1	1185	Db	69236	ACCGAGTTCGCGCGCGCGCTCGCGCTCACCGCGCACGGGAGCGGTGTGTTCAC	69295
Db	68292	GCGCACCGCGAGATCAACCGCGCGCGCTACGCGGAGCGG-----CCTTCTCTCTTACCTC	68348	QY	1471	sAspValArgSerLeuIleGlySerProGly-----ArgThrPheProProValHisPr	1489
QY	1185	yValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProG1	1205	Db	69296	CACGCGCGCTTCTTCCACGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	69355
Db	68349	ATCGCGCGCATGACGCGATACAC-----CGCTCTACGCGATCTCT-----CCTGG	68396	QY	1489	oLeuAspValMetAlaAspAlaArgAlaLeuGlu-----	1500
QY	1205	ygIleThrLysGlyIleProSerThrArgVal-----	1217				

[illegible]

Qy	1756	pArgLeuAlaTyLeuProThrAlaProGlnProPhe---	SerSerArgHisSerSerSe	1775
Db	70386	----- ---CCGCTTACCACCGCTCGACCTCGCGC	70414	
Qy	1775	rProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerG1	1795	
Db	70415	ACCACC-TGGGACTCGGGCCCGCGCGCGTTCACCGACCGCTCGTCTTCGGCGCC	70473	
Qy	1795	uArgGluArgAspArgAspArgGluArgAspArgGluArgGluLysSerIleLe	1815	
Db	70474	GACCTCGCGCAGCGC-----CGCGCCCTCC	70500	
Qy	1815	uThrSerThrThrThrValGluHisAlaPro---lleTrpArgProGly---	1833	
Db	70501	TTCCGGGAGCTACGCTCGCGCCCGAGTCCGGGTTCTTGGACGCGCTCGTGCACCTACCTC	70560	
Qy	1833	nSerSerGlySerSerGlySerSerGlyGlyGlyGlySerSerSerArgProAlaSe	1853	
Db	70561	CCCTACCGGCGAGCTCGTGGAACTCGCAGGGAACCTGGGCGCGTCAACCGCCCGCGCA--	70618	
Qy	1853	rHisSerHisAlaHisGlnHisSerProIleSerProArgThrGln-----As	1869	
Db	70619	-----CCGCGCGCACTGGGAGTGGCGCTGAAC	70647	
Qy	1869	pAlaLeuGlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleleThrAl	1889	
Db	70648	TTCTGCCGCAACCCGCCCA-----	70666	
Qy	1889	aValGluProSerLysProThrValLeuArgSerThrSer-----Th	1903	
Db	70667	-----CCAGCGCCCAACCGCGCGCAACACACCTCGCGAAACGGCGCTGTCATC	70719	
Qy	1903	rSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyG1	1923	
Db	70720	GAGCTGTTCGGAGAGCCGACCTGCTCGCGCGCGCGGACCGGTCCCGCGCACCGGTGG	70779	
Qy	1923	YThrLeuAspGlyValTyProThrLeuMetGluProValLeuLeuProLysGluAlaPr	1943	
Db	70780	GACGG--CACGGTGCTCGCCCT-----CTCCCTAGCGCAACTCG	70816	
Qy	1943	oArgValAlaArgPro-----GluArgProArgAlaAspThrGlyHisAlaPheLe	1960	
Db	70817	GCAGACACACCGTGCTGGTCTCTCGACGCGACCGCGA-----	70853	
Qy	1960	uAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerG1	1980	
Db	70854	-----CCACCGCACACACGAAACCGCGACCGGCTGCTCACCGGATGACGAGAACGC	70906	
Qy	1980	uPro-----ArgProLeuValProProValSerGlyHisAlaThrIleAlaAr	1996	
Db	70907	TCCTGGCGGCGCTCGCGACCGGAGCCGCCCT-----GCCCGCTTGCCTCG	70954	
Qy	1996	gThrProAlaLysAsnLeuAlaProHisAlaSerProAspProProAlaProAl	2016	
Db	70955	CCCCCGCGCA-----CACACGAGGAGCCACCGATGACACGACCCCGCG	70999	
Qy	2016	aSerAlaSerAppPro-----HisArgGluLysThrGlnSerLysProPheSerIle--	2033	
Db	71000	GACCGCGCGGACCCACCTACACAGCTGGTGTCAACGACGAGGAGTAGTACTCGATCTG	71059	
Qy	2034	-----GlnGluLeuGluArgSerLeuGlyTyrHisGlySerSerTySerPr	2050	
Db	71060	GCTCGCGGACAGGATCCCGGC--CGGCTGGCGGC-----CACCGGAACCTCCCGCACCC	71115	
Qy	2050	oGluGlyValGluProValSerProValSerSerProSerLeuThrHisAspLysGlyLe	2070	
Db	71116	AGGAGGAGTG-----CCTCGCGCCACATCGACGAGGTGT	71148	
Qy	2070	uProLysHisLeu---GluGluLeuAspLysSerHisLeuGluGlyGluLeuArg-----	2087	
Db	71149	GGACGACATGGCCCCCGCCGACCTCGCGGAGGACCTCGCGCGGAGCAGCGCGAGC	71208	

QY	2088	-ProLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuAr	2107
DB	71209	CGGCTCCGCGCCCGCGCCCGCGGA---GGAGGAGCCGAGCCTCGTCGACGGGCTCTGGC	71265
QY	2107	gProLeuProGlu-----SerGlnProSerSerSerProLeuLeuGlnThrAlaPr	2124
DB	71266	CGGGCAGCAGCGCGGTGGAGTCCGTCTCCGCGCGAGCGACGCGCGCGCGCTCGCGG	71325
QY	2124	oGlyValLysGlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleTh	2144
DB	71326	AGGCGGTGACCGCGCTACTGTTCTGTCGCTTCGCGCCACCGCGCGCGGACCGAAC	71385
QY	2144	xGln-----AspTyrThrArgHis-----	2150
DB	71386	TCGGCGTTCGCGTCGACCCCGCGGCGACCATCGAGCGGACCGAGTGGCGCTGACCG	71445
QY	2151	-----HisProGlnGlnLeuSerAlaProLeuProAlaPro-----Le	2163
DB	71446	GCACCTCAACCTCGATTCGAACCGGTGCGTGCACGCGCGGTCTGACGTGACCACT	71505
QY	2163	uTyrSerPhe-----ProGlyAlaSerCysProValLeuAspLeuArgArgProPro--	2180
DB	71506	TCACGGCGAGGCGCGCTGGAGCGCG-TGTCCGGACCTGACCCCGCGCGCACCGCG	71564
QY	2181	-----SerAspLeuTyrLeuProProAspHisGlyAlaProAlaArgGlySerPr	2198
DB	71565	CCGTGAGCGCGGTTCGGACCGGCGCGCGACCCAC-----CGAAGGAGGGAC	71615
QY	2198	oHisSerGlu-----	2201
DB	71616	CCATGACACCCCATGACACCCACGACCGACACCAACCGCGACCGCGCTCT	71675
QY	2202	-----GlyGlyLysArgSerProGluPr	2209
DB	71676	TCGCCACCTCCGCGCCCGCGGCTCGGGACCTCTCCAGCGCAACATCGGCTCTGCC	71735
QY	2209	oAsnLysThrSerValLeuGly---GlyGlyGluAspGlyIleGlu-----ProValSe	2226
DB	71736	TCGTCCCGCGCGCGCGCGCGAGCGGTACACCTGTCTCGCGGAGGACCTGGCGG	71795
QY	2226	rProProGluGlyMetThrGluProGlyHisSerArg-SerAlaValTyrProLeuLeuT	2246
DB	71796	CCCGTTCCGTCCGGCATCACCGCCACACGTACCCACCGAGTGTGCTGCCCTGC---	71851
QY	2246	yrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrS	2266
DB	71852	-----CCCGAGCGG---GGCGAGCGCCGACCC-----	71875
QY	2266	erGlnProProAlaPheSerLysLeuThrGluSerAsnSerAlaMet-----ValL	2284
DB	71876	--CGGTGCGCGCTTCCTCGGACCTTCGCGCGCGCGCGCTTCGCGCTTCGCGCTCG	71933
QY	2284	ysSerLysGlyGlnGluIleAsnLysLysLeuAsnThrHisAsnArg-----AsnGluP	2302
DB	71934	ACCCGAGACGACGAGCGCTGCACGCC-----GGCCACGCGCGCGCGCGCGGTGC	71984
QY	2302	roGlu-----TyrAsnIleSerGlnProGlyThrGluIlePheAsnMetPro-----	2317
DB	71985	CCGAGCGGATCGGCTCGCGAGACCGCGCGGAGACGACATACCCATCCATCC	72044
QY	2318	-----AlaIleThrGlyThr---GlyLeuMetThrTyrArgSerGlnAlaValG	2333
DB	72045	GCCTCCACGTCCTGTGGGGGACCGCGGACCTGTACGAGTAGTCGCGCATGCGC---	72097
QY	2333	lnGluHiAlaSerThrAsnMetGlyLeuAlaIleIleArgLysAlaLeuMetGlyL	2353
DB	72098	-----CTGCGCGCGCTGGCGCTTCGCGCACCGCGCGCGCGCGC-----	72139
QY	2353	ysTyrAspGlnTrpGluGlu-SerProProLeuSerAlaAsnAlaPheAsn-----	2369
DB	72140	-----GACGTCTCGGAGTGTCGCCCGCACCGCGCGCGTCCGCGCGCGCGCGT	72194
QY	2370	-----ProLeuAsnAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGly	2388

Db	72195	TGCCCGCTCGTTCGTCGCGTCCACCCGCGGGGCAACCGACCTGGACAGGATGGC					72254
Qy	2389	ArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGlyArg					2408
Db	72255	CGCTCGAGC-----ACTAGCGCGCTCTCGCGCGCTCGCGCCGAACTCTCGG					72305
Qy	2409	ProSerSer-----ArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAs					2425
Db	72306	CCTCCCTCTGCTGCTGGGCGCAAGACCGACCGCCGAGCTGGAACTGCTCCGGCAGC					72365
Qy	2425	pArgProProSerValSerValHisSerGluGlyAspCysAsnArgArgThrProLe					2445
Db	72366	CCGTCCTTGACCGCGTCCCGCG-----ACCGCTCGTCCACCTCG					72404
Qy	2445	uThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPro-----					2461
Db	72405	AGCGCGCGCGG-ACCTGACCGACCGCAACGCTCTCGCGACCGCAGCTGCTCGTC					72463
Qy	2462	-----PheProTyraSnProLeuLeuMetArgLeuGlnAlaGlyValMetAlaSerPr					2479
Db	72464	GCCAACGACTCTCGCTCGCGCC-----ACGTC					72490
Qy	2479	oProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisAlaTrpAs					2499
Db	72491	GCCGCGCGCTCCGACCCCGCTCGTCTCTACGGC-----					72530
Qy	2499	pGluGluProLysProLeuLeuCysSerGlnTyrgluThr					2512
Db	72531	----CGACCGGCACCGAGTAGCTGTCGACGAGATCTACC					72566
RESULT 88							
LOCUS	AE007093	14860 bp	DNA	linear	BCT 27-APR-2001		
DEFINITION		Mycobacterium tuberculosis CDC1551, section 179 of 280 of the complete genome.					
ACCESSION	AE007093	AE000516					
VERSION	AE007093.1	GI:13882291					
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
FEATURES							
source							
gene							
CDS							

QY 1131 lGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetG 1151
Db 4968 -----CACCAGCCCGCCGCGCCACCA-GCCCCACCGTTGCCCGCCAGCAACGAGT 5019
QY 1151 yLeuPro---LeuProMetAspProLysLysLeuAlaProPheSerSerGlyValLysGlnG 1170
Db 5020 CCCTCCGAGGACCGTGGCGCGCGTCCGCCACCTTGGCGCGCTCA----- 5065
QY 1170 uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaG 1190
Db 5066 -CGCGCGTACCGATGTGGCGCGCTCACCGCC-----TTGCCTCAGCCGCC 5112
QY 1190 nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLy 1210
Db 5113 ACCGGCCCC-GCCATCACCGCCAGAGCTTCGGCAGCGGTGGCGTCCGCCATCACAC 5171
QY 1210 sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh 1230
Db 5172 CGGCTCCGCGTCCCGATAGCCCGCGTTCGCCCGTCCACCGCTGGCGCGCGTGC 5231
QY 1230 rHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleIleGlyLus 1250
Db 5232 CGCGCGCGCGCGGC-----CTTGGCACCGTTGCC----- 5262
QY 1250 pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy 1270
Db 5263 -GCCAGCGCGCGTCCCGCGTTCAGTGGCGCGGTGGTGGCGGTCCACCGCT-- 5319
QY 1270 rGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCys 1290
Db 5320 ----GGCGCGGAGGTGCGTTCGGCTGTACCCCGCACCGCGCGACCGCTGCGC 5375
QY 1290 rLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgTh 1310
Db 5376 C-----GGGTCACCGCGCTTACCGCGCGCTC 5402
QY 1310 rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe 1330
Db 5403 CGCGCGCGCGCGCGATGATACCAATGCCCGCGT-----GCCGCGCGCCCGCCACGC 5459
QY 1330 uMetGlyArgAlaIleProGluArgHisSerProHisHisLeuLysGluGlnHisH 1350
Db 5460 CGCGCGCGCGCGA-GTCCGCGCTGCCACCGACCCACCGTGGCGTACTGCCGA-- 5516
QY 1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnAspTy 1370
Db 5517 -----CCGTCCCGAAGGTGCGCGCGCGCGACCCCA 5548
QY 1370 rLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProPro 1390
Db 5549 CCGTCCCGCGCATGTCACCGACCCCGCGGACCG-CCATCGCGCGCACACCGCGCGC 5607
QY 1390 rArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAl 1410
Db 5608 CCCACCGTGGCGCGGATGTCGATA---CCGTGGCGCGCGT----- 5650
QY 1410 aHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProAr 1430
Db 5651 -----GCCCGCGC 5658
QY 1430 gGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGly-- 1449
Db 5659 CCAACCGTAGCGCCACACCGCGTTCGCCCGCGCGCG-CCGTGGC-GCCGCGAC 5714
QY 1450 -SerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLy 1469
Db 5715 CGCGGTACCGACACCGACCC-----ACCGGCGCCAC----- 5748
QY 1469 sLysHisAspValArgSerLeuIleGlySerProGlyArgThr-PheProProValHisP 1489
Db 5749 -----GGCACACCGGACACCGCGCGCGCGCGCGCGCTGCC 5783

QY 1489 roLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuL 1509
Db 5784 CC----- 5785
QY 1509 ySerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgGlyAlaProValI 1529
Db 5786 --GCGTGGCGCTGACCCCGCTTCCCTCCCGTGGCCGAGGCGAGGGCATCT- 5836
QY 1529 leValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaP 1549
Db 5837 -----GAACAGGCGCGCCCGAATTGGCCCGCTTGGCGCGCGCGCCCATCGCCAC 5888
QY 1549 roPheAlaGlyHisLeuProArgGlySerProValThrMetArg-----GluProT 1566
Db 5889 CA-----TTGCCGCGCGCGCCACCATCGCGACCGCGCGCATTCGCGCGCT 5936
QY 1566 hrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuT 1586
Db 5937 CGCCTCCGTGGCC-----C 5951
QY 1586 hrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProH 1606
Db 5952 CGCGCGCGCGCGAGCTGGCTTGGCACCGTGGCCACCGTTACACCGTTGCCCGCGC 6011
QY 1606 isProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgS 1626
Db 6012 TGCGG--GGCCGCTGG-----TCGCTGGCGTCCACACCGCTG--GCCG 6050
QY 1626 eHisIleProLeuAlaPheAspProThrSerIlePro-ArgGlyIleProLeuAspAla 1645
Db 6051 CGCGGTGGCTTGGCGTGGTCCGCCCAAG--CCGCCGAGGCTCCCGCGCGCGGGT 6107
QY 1646 AlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyr 1665
Db 6108 CACCGCGCTACCGCGTCCCGCCCATCA-CTGCCATGACCGCGCTCACCGCGCTTGGCG 6166
QY 1666 Pro-----ProTyrLeuLeuArgGlyTyrProAspThrAlaAlaLeuGluAsn 1681
Db 6167 CGCGCTCCCGCGAGCCCGGC-GTCAACCGCAACCGCGCGCACCGCTGCT----- 6216
QY 1682 ArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisThrAlaThr 1701
Db 6216 ----- 6216
QY 1702 AlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAla 1721
Db 6217 -----GCCGACCTGACCGCGGGTGGCGCGCGCGCGCGCGCATC--- 6258
QY 1722 LeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuPro 1741
Db 6259 -----ACGCGCGCGCGC----- 6270
QY 1742 valLeuVal-ProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLe 1761
Db 6271 --GCCGTACCGCCACCGCGCGCGCACCGCGCTCGCG----- 6307
QY 1761 uProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyG 1781
Db 6308 -CTTTCCCGCGCGCGCTCCACGGCTCAGCGCTCGCGCGCGCGAGCGCGC----- 6361
QY 1781 yProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgAs 1801
Db 6362 -----TTGCCGCGCGCGC-ACCACTAGCTCCGCG----- 6390
QY 1801 pArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThrVa 1821
Db 6391 -----ACCACCGCACCGCGCGCGCGCGCTTGGCCACCGGTCCACCGTCCG 6437
QY 1821 lGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSerGlySerSe 1841
Db 6438 CCACACCGACCC-----ACCGCGCGCACCGCGCTCCGC 6470
QY 1841 rGlyGlyGlyGlySerSerSerArgProAlaSerHis----- 1854

RESULT 89
 APO05045/c
 LOCUS
 DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section 25/30.
 ACCESSION APO05045
 VERSION BA000030
 KEYWORDS APO05045.1 GI:29609643
 SOURCE Streptomyces avermitilis MA-4680
 ORGANISM Streptomyces avermitilis MA-4680
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
 Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
 Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
 21477403
 11572948

2 Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.
 Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
 Nat. Biotechnol. 21 (5), 526-531 (2003)
 22608306
 12692562

3 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
 Direct Submission
 Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
 (E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)

This work was done in collaboration with Haruo Ikeda (*1), Jun Ishikawa (*2), Akihiro Hanamoto (*3), Chigusa Takahashi (*3), Mayumi Shinose (*3), Hiroshi Horikawa (*4), Hidekazu Nakazawa (*4), Tomomi Osonoe (*4), Norihiro Kishida (*4), Hisashi Kikuchi (*4), Tadayoshi Shiba (*5), Yoshiyuki Sakaki (*6, *7), Masahira Hattori (*1, *7) and Satoshi Omura (*1, *3).
 Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa

*1 Kitasato Institute for Life Sciences, Kitasato University
 *2 National Institute of Infectious Diseases
 *3 The Kitasato Institute
 *4 National Institute of Technology and Evaluation
 *5 School of Science, Kitasato University
 *6 Institute of Medical Science, University of Tokyo
 *7 RIKEN, Genomic Sciences Center

Following url is also available.
 http://avermitilis.lis.kitasato-u.ac.jp.
 Location/Qualifiers
 1. 299925
 /organism="Streptomyces avermitilis MA-4680"
 /mol_type="genomic DNA"
 /strain="MA-4680"
 /db_xref="taxon:227882"
 /note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"
 140..1522
 /gene="amyA4"
 140..1522
 /gene="amyA4"
 /note="SAV5981"

gene
 CDS

```

/codon_start=1
/transl_table=11
/product="putative alpha-amylase"
/protein_id="BAC73693.1"
/db_xref="GI:29609644"
/translation="MARKALAVSLAALAGLTVPPAVSPPTTAAASPPTKDVAVLFEW
KFDVSARECTNLTGAGYVQVSPGPAHIGSQSWTTSYQVSYKISAGRLGDAFAQN
MINTCHTAGVRVVDVTVMHMSAGSGTGGSAITKYNPGLSYSSYSGRLGDAFAQN
TRANVONCELVGLADLTGEEYVKRTIAGYMTLLGYGADGPDVAVKHHPAALAN
IKSLTNPVYKQEVYIYASGEAVQPTGYTGVQDQYERAYDKRVFNENLAVLKN
YEGWGLNSVAGVVDNHDTERNGSTLNYKDGANTYLANFVLMPLAYPGAPDINSY
EWSADAGAPGGGTACWQGWKQHAWEIKAWAFRNPATRGVSNVNTGVTSSGRTTLAGTA
AFGRGAGYVAINHESGLRTRYQTLTAGTYCNVQNTGTVDSSGRTTLAGTA
LALYSGKSTC"
1661..7081
/gene="amyX"
1661..7081
/gene="amyX"
/note="SAV5982"
/codon_start=1
/transl_table=11
/product="putative pullulanase"
/protein_id="BAC73694.1"
/db_xref="GI:29609645"
/translation="MDLIPRGSAPRTRRHSRTAAAVAAALAVSLAQPLEARAAT
PAPPSDAKLAAPARHDATREQYFVMPDRFANGDTSNDKGLTSGSLSTGYDPTDK
PFYQGGDLGLTRKLDYIKGLTTSIMLAPIFKQNPQVQGTGKDGASGYHGYITDPTK
VDHPGTWKDLLETILSKAHAGMKVFFDVTNHTADVDEKSYGLSKGAPVYLTK
DGRPDDAGYTDGPRKPAVDGDSFPRTPAARAKNAKVPKSLNDPTMHNRGSTF
AGESSTHGDGSLDILWTERPEVVRGMEKIYEKVRDGFIDGFRIDTIVKHNTERPTQ
WATALDAYAKRGKDDPFMFGEVYSADTSVTSVYVQGRDSTLIDFPFQDARSVASQ
CGSAAKLASVFGDDYKTTDKANAYEQVTFILGNDHMGRIYGLFNDNPKATDAELLRK
DLANELMFLSRNPVYVYQGGTGGGQDKARQTMFASKVADYLDDEIGTDRGH
ASDADTSAPLYKEIATLKLKONPALADGIOTERYAADGAVAFRTDARTSTEY
VYAVNNAKASATFATGSADTAFKGIHTDDVLKSDADKKTIVTPAGAAVLAAG
RGTTPARKPSLTLPKAGATVLSADVDGRLNRVVFQAQVNAKWEULGSADHA
PRYRTTKGVDPAGTALRYKAVIDAAGRATASATSTGTPPAETPTASSRDYAI
VHYKPDGDDYTDRLYAWDLGDSSTVPGHDFVGRDAYGAFVAKLPGASVNF
LVIDKDGKDSADRTIDVTKAGEVYVQGGTETVTRPDYPAQDKTAVIHYHRADG
DLTGMLHWMTGAATPDWSKPLEPVDYAGAVEPLTDGKATDLSYIHKDSEKDL
SADRSLDTADGHEVNLGONGHLLPOPAGSAAALDLTTSKAVINDRNTVANGSDA
AASQTLSDRSGSIANKGSLTSDDERWLRLSKTSLTDAQKAAFPHLSKYTASVDP
DDRVREALAGVQASQAANGAVLAATGVLQAGVLDLDDYKPYRVKVAAPTVRKY
AWYATTAQSVSLDGAHVNRKNATGVSVTPASMKGKPYRVKVAAPTVRKY
TNKVTDPYVALTIDRSVLDLDRSLAPSGNSLKKPKAPLRLDAEIOELHTRDF
SVADRTVPAKDRGTYLAFTDKNSDGRSLROLAESGTSYVHLLPAFDIATIAEKSGQ
QATCDLASAADSEKQOECLTAAKADAVNMGYDPHYVTPEGSVATDANGTRTVE
FRMWKSLNQDGLRVVMDVVVNHATAAGAGTSVLDLRIVPGYVQRLADGSAVATCC
ANTANAMMGLVLDVSLVTWAKYKVDFGRFDMGHQPKANILAVRKALDALTYAKD
GVGKKIILYGEWNFGEVADDAFVQATQKNMAGTATFSDRDARDAVRGGPPDAD
PGVQGFSGLYTDPNSDANGTPAQKARLLHYODLTKVLSGNLAKRYFTDSSGKEV
TGSEVDYNGTGAGYADAPGALAYADAHNLSYDALTYKLPKGTAGDRARMQVLAM
ATYALAAQGSLSQAQSDLLRSKLDNSYDSGDWFAIHWNCODGNGFGRLPMAADN
KSKMPYATPLISVVKVGCDOIGTSAGYODLIRITTEPESLSTAGOVQSKLTPLS
GKDETTPGVITMKLGLVFNFAVNPDOQEQVLAALAGKVALHPVQAGADPIKVSAY
TAKSGMFAVPGRTVAIFSQAR"
7152..8756
/note="SAV5983"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC73695.1"
/db_xref="GI:29609646"
/translation="MDANGRTQDSTVWVDDVPTSTYRGMAGVLRAGHSVVPVASAGE
ALVELDLVLRAGSLPDVALVDGLPDMSGFELCRKKSQPTPAALPVHFFSAALAPG
DRQGLDAGDAYLSVPAPERIEOAVVRAVARGARLRTGAELARELTLSETIVAVQ
AKRCPTELAHAAQAGARTITGDFAAVFLGEGELYPGLSRRAPSLPQGAHEAVA
RLIRGTGAGVRSIVVPAPLPAFLPAGFLPQSGEDARVLAQTOQEGKPAVCLTPAHA
PAQETELVARLAATALAAEPLMFLYQVERHVALTLQHSFLPKPKADLPQMEVAVRIV
RASQTEIGDGFYATLRSGVLTGVGVVHSDLAATVWIRHAYRAGVAGVDDPPA
VLVTRIDRLQHYHPEATVCLALVDPDSGHVSIANAGHIPPLIYVHDDGTARYPGA
GPLLGLRPPATEFLTPRHLMLVTDGLITETCTGTDLSVSNQELRMAAADPSPGVG
ALCDALLDCFGDRDREDDIALALRLG"

```

CDS


```
CDS
8788..9144
/note="SAV5984"
/codon_start=1
/transl_table=11
/product="putative isomerase"
/protein_id="BAC73696.1"
/db_xref="GI:29609647"
/translation="MPQITVDYSAEALAPGDFRPFARALHEEVVTRTAAKPEACKTOF
RGEDTTVGADGEGHAIHVHTLGLLAGRSDETKARLTEAALLRLQYVKPDGGGLAHA
SAEVRDLDPSRKFDPA"
complement(9148..9849)
/note="SAV5985"
/codon_start=1
/transl_table=11
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC73697.1"
/db_xref="GI:29609648"
/translation="MTTGVRRMGVEERRQOLIGVALDLPSRRSPDEVSIDETASAAG
ISPLVTHYFGKLSLTHAALQASDDLAGFVPEFGPLGARLLRVRRFFDFVEEH
GPFSAIMRGGAAGVSSSTNALVDGVQAAVQVILSHLKVTEPPVRLVLRWISLV
ESTALIWLDRRIARGSEVQLVHDFAAALAAISAAYDDELAALLRDLRDEAPDGPFS
DLVRLIALTSPPTA"
complement(9928..11844)
/note="SAV5986"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC73698.1"
/db_xref="GI:29609649"
/translation="MQSQDVAPRPTPASGPPVVEQAEALVHHYPLRLVLYLVLP
SLGNRVLTAHALTQALPRGRTSAVIPAQAKGRDQDPGLVLRVLRVLTALVLEAGL
PLGMWPKRSQPLPLQVWGLRSLFRRSGADELDALDLSLSPARAAHVLRGLE
KLSDSVRRVLDIAGVADPAATAEAIAKVPAPDALDPSFPCSLHARPTDLRRRQ
HMAALAAAVAVCGALIGLPGGWAPDGAAPSAVAENPAARALDPARLTKTAATA
WKTSAFTDFSWPARGDRTGDTALLRELAIVAPRGEVRSATPGTSPGAPGPOL
LYAGDVSAHVLLYDGLRIARYAEPKDTGAALDFARVDGSAEAREALVLRGAD
NRYLTAPWVYAAERDLMPASGMOLNLDGVAAPLAGSAQRTGATSNALQLTD
GSGRLTLDLGLTIPAHLTTPSPAPREASRTALHSMWPFACSLASVRSQGVSVNA
WQQAQALPDGSGSAGVCTRADTWRGGTRVLAQFHTPGGAYGAVAAKAEDSPACGA
RDHVLAVLWKSQGTAWYLLAAGSSDTSIRATGGIDGAEGNLLVTRTERGARAGL
KGVLDGRPTISGLR"
complement(12002..12514)
/note="SAV5987"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC73699.1"
/db_xref="GI:29609650"
/translation="MGFPVTENNWDERVAWAATPDDYAEKDAADFRAVVDALVAELPA
DSPFPPEQACAMDSHGSDKAVPLYREALARGLSYGRRAKTLQSSSLRNTGQAE
GVKLLTPELGPSDELDDAVRACLALCLADLGRERGLSLVLGALAPHLPRYQSRMAN
YARLLVPEPD"
complement(12571..15687)
/note="SAV5988"
/codon_start=1
/transl_table=11
/product="putative alpha-N-acetylglucosaminidase"

Alignment Scores:
Pred. No.: 0.000253 Length: 299925
Score: 665.00 Matches: 640
Percent Similarity: 30.98% Conservative: 255
Best Local Similarity: 22.15% Mismatches: 1090
Query Match: 5.03% Indels: 918
DB: 1 Gaps: 126

US-09-522-753-5 (1-2517) x AP005045 (1-299925)
Qy 2 serGlySerLeuValAlaGlnThrTrpArgAlaThrGluPro-ArgTyrProPr 21
Db 278628 TCAGCAGTAGTACGGGACTCGCGCTCGACCTGCGCGCCCTCTTCGCGCGGACACCTCG 278569
Qy 21 oHisSerLeuSerTyrPro-ValGlnIleAlaArgThrHisThrAspValGlyLeuLeuG 41
```

```
Db 278568 GCACACGCATCGGCGTCCGGATCTCGAGGCTCCGCGAGCGGTCTCGCGCACCATGTC 278509
Qy 41 luTyrGlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleLeuGlnP 61
Db 278508 CCCTCGAGGCAACACCCAGCCGTACGGG-----CTGTCTACGCGCGCGCGC-CTCCGCC 278456
Qy 61 roGlnArgArgProSerLeu-----LeuSerGluPheGlnProG 75
Db 278455 GTCTCTCGCCGAGACCTCGGCTCGGTCTCCATGCTCCACGCGCGGCGAGCTCCAAGCTC 278396
Qy 75 lyAsnGluArgSerGlnGluLeuHisLeuArgPro----- 86
Db 278395 GCCGTGGCGCTCGACCTGCACTGCACGCACACCGCGGGCGTACGTCGCGGCTCGTCAAC 278336
Qy 87 --GluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPheIleGluSerL 106
Db 278335 GGAGTGGCCACACCCGTGACCGGGGGCGTCCACGGGAGCTACGAGATCGTATCAGC 278276
Qy 106 ysArg-----ProArgLeuGluLeuLeuProAsp---ProLeuLeuArg---ProS 121
Db 278275 GACGAGGAGGCAAGCGGTGTGCACCGCGCTGACCTGCTGCTGCGCGAACTCCCC 278216
Qy 121 erProLeuLeuAla-----ThrGlyGln----- 128
Db 278215 GCGAATAGCGCGCTTCGGGGCGCGGACCGGCGGGAAGGACGGGACCGGCG 278156
Qy 129 --ProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyLysLeuGluProV 148
Db 278155 TACCGCGGGAACCGGACCTTTGGCCCGACA-CGAGATCATACCGGC---CTCGGGCCCG 278100
Qy 148 alserProSerProProHisThrAspProGluLeuGluLeuValProProArgLeuS 168
Db 278099 TCGAACCGGACTTCCACCGCCCGCGAGCC-----CTCACCCCGCCCGCAGCA 278049
Qy 168 erLysGluGluLeuIleGlnAsnMetAspArgVal---AspArgGluIleThrMetValG 187
Db 278048 CGCGCTCGGGG-----CGCAACTCCCGCGCTGCGCGGACCGTTGGGCCACACCGCCCG 277995
Qy 187 luGlnGlnIleSerLysLysLysLysGlnGlnGluGluGluAlaIalysP 207
Db 277994 TCCACACCGCGCGGCGCTCCAGCGCTCCCGCGCGCGCTG---GCCAGCGCGCGCGCC 277938
Qy 207 roProGluProGluLysProValserProProIleGluSerLysHisArgSerLeuV 227
Db 277937 CCGCCCGCGCGCGCGACCTTACCAGC-----TCTCGCCACCGCGCAGCACA 277893
Qy 227 alGlnIleIleTyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIle----- 244
Db 277892 CCCTCCCGCGCACTGACAGCTCAGCGCGCGCGCGCGCGCTCGCGCGCGCATCTCC 277833
Qy 245 -----LeuGluGlyLeuGlyProGln-ValGluLeuProLeuTyrAsnGlnProSer 261
Db 277832 GTACCCGACTCGCGGCGCTGCCCGCCACGCGGTCAACATCCCTC----- 277791
Qy 262 AspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeuIle 281
Db 277790 -----CGCCTGGACCTGATCTCGCGCACCCCAAGCAGCAGCAGCGCTACAGCTTC 277740
Qy 282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys----- 299
Db 277739 CTGCTCGCGCGCAACCGCCACCGCACCTCTCTCCCGGTGCGCGCACCGCGCTGCCCGCA 277680
Qy 300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsn 319
Db 277679 CAGGAAAGTAACTGTGTAACTGTCATGCAATCAAGG-----AAT 277638
Qy 320 ProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGlu 339
Db 277637 GCGCGTCCGCGCGCTTGGCCACACGCTC-----CCTCAGACCGCGCGC 277596
Qy 340 IleArgLysGlnArgGluLeuGlnGlnArgMetGlnSerArgValGlyGlnArgGlySer 359
```

Db 277595 CCAGCCCGCCCGCCCTCGAAGATCACTTGGAGCCC-----CTGAGCGGAACA 277545
Qy 360 GlyLeuSerMetSerAlaAlaAraSerGluHisGluValSerGluIleLeuAraGlyLeu 379
Db 277544 GGGCGTCTCACCATGGGT-CACGGGAGCAGCGCGCCAGGCCCCCGACCGACCGCCC- 277487
Qy 380 SerGluGlnGlnLeuGluGlnMetArgGlnLeuAlaValIleProProMetLeu 399
Db 277486 CCCACCCCGGACCAACCCAGCGCATACAAAGAGCGTCACAGCTGTCTCCCATCTCCTC 277427
Qy 400 TyrAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspPro 419
Db 277426 CCCGAGCCC-----
Qy 420 MetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr 439
Db 277417 -----GCCACCTGGCGCTTATAGTCAACGGCCAGTCACCGCGC 277382
Qy 440 PheArgGluLysPheMetGlnHisPro-----LysAsnPhe 451
Db 277381 CGCGGGCGCGTCTGACAGCAGCGCCAGCAACCCAGTACGGCCCGCGGAGACACAG 277322
Qy 452 GlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyr 471
Db 277321 GCCCTCCAGCAGGAGGTGCGCCAGGAGGAGTGTATGTCGACACCGTTCTTGCT 277262
Qy 472 LeuThrLysLysAsnGluAsnTyrLysSerLeu-----ValArgSerTyr 487
Db 277261 GATACTCACCACCGTGTCAACCGGAGCACTCACGCTCACCGCTTTCGGATCGCGCGA 277202
Qy 488 ArgArgArgLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 507
Db 277201 CGAGCAAGGGCAGCAGCAGCGGACAGACACCGTGTATCGGCTCGAGCGGCC 277142
Qy 508 GlnGlnGlnPro-----MetProArgSerGlnGlnGluLysAspGluLys 523
Db 277141 CCTCAGCGCCAGAACTCCGCCCGCGCTCGGATCCAGTACGGCTCCAGATCGCGCT 277082
Qy 524 GluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGlu 543
Db 277081 CGACGA-----CGCAACAGAAAGAACTCGTGCCTCGCGCGCT- 277046
Qy 544 AspLeuLeuLysGluLysThrAspThrSerGlyGluAspAsn-AspGluLysGluAl 563
Db 277045 -----CACGTTCAAGTCAAGCCCTCGACGACAG--GC 277013
Qy 563 aValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArgArgLysGlyArgIleTh 583
Db 277012 CGTCCCGCCAGCGCCAGCAGACCGCCAGCGCTCGTCCCAACAGGACGCTCTC- 276955
Qy 583 rArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaG1 603
Db 276954 -----GGCGTCTGTCGCGCCGCTGAACTCGGTTGCGCCACCCAGATGACGAGGCTTCGA 276899
Qy 603 uLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluGluMetGluThrAl 623
Db 276898 CACGCCCGGCTCTCGATCTCCCGTCGACACCCCGCCCGAG----- 276853
Qy 623 aLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMetValGlySe 643
Db 276852 -----CTGACCCAGGCAAGAACTGGCAGAGC----- 276826
Qy 643 rLysThrValSerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAs 663
Db 276825 -----AAGAGGAACG-TCCTTACA 276807
Qy 663 pGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgLysLys 683
Db 276806 AGACGTACTTCGCGCCCGCCACCGAGCGCTCCAGGGCGGTTCGCGCGCGACTACG 276747
Qy 683 sLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValValGluAspG1 703
Db 276746 CCTACAACACCT----- 276734

Qy 703 uGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAl 723
Db 276733 -----GAAGAAGCGCAAGG 276720
Qy 723 aLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrValAs 743
Db 276719 TCTTCATCG-TCGACGACAAGCAGACCT-----ACGGCGCGCGCTCCCAAG 276673
Qy 743 nAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThrG1 763
Db 276672 CTCTTCAAGCAGAACTTCAAAA-----AGCGCGCGCGCAAG- 276636
Qy 763 yGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGlyProProProGlyProPr 783
Db 276635 -----TACCGCGCGAGGACCATCATCAGCGTCGGCGACACGACTTCTCTCG 276589
Qy 783 oThrProProArgArgThrSerArgAlaPro----- 793
Db 276588 CTGCTCACCAAGTCAAGAACTCCGGCGCGACCTTCTACTACGGCGCGCAGTACGAC 276529
Qy 793 ----- 793
Db 276528 GAGTCCGAGAGATCAACAGCAGCTCAAGGACGGCGCGCTCGACATCCCGTGTTCGCG 276469
Qy 794 -----IleGluProThrProAlaSerGluAlaThrGlyAlaPro----- 806
Db 276468 GGCAGCGGCATGTTCTCCGACACCTATCATCGAGCCGCGCAAGCCCTCCGAAGCGCAG 276409
Qy 807 ----ThrProProAlaProProSerProSerAlaProProPro-----ValVa 822
Db 276408 CTGCTCACCTCGCTCGCGCTCCCGTCAACCCCTCCCGCGCGCCCAAGAGTTCATCGCC 276349
Qy 822 lProLysGluGluLysGluGluThrAlaAlaAlaProProValGluGluGluG1 842
Db 276348 ACCTCAAGCGGAGAGTACCCCGCGACTACGGCACTTACGGTGGCTACTCTCTACGAC 276289
Qy 842 uGlnLysProAlaAlaGluLeuAlaValAspThrGlyLysAlaGluGluProVa 862
Db 276288 GCGCCACCGCCATCATCAAGCCATCGCGCAACGCTCGTGAAGCAGCGCAAGTCCCT- 276231
Qy 862 lLysSerGluCysThrGluGluAlaGluGluGluGluGluGlyLysAspAlaGluAl 882
Db 276230 -----CGACGCGCGCA----- 276219
Qy 882 aAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGlySerGlyAr 902
Db 276218 -----CCAGATCTCGACGAGGTCAG 276196
Qy 902 gAlaThrThrAlaLysSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSe 922
Db 276195 AAGACCAAGTTCGACGCGCATCGCGCGCTTCGCGCTTCGACGAGTACGCGCACA----- 276141
Qy 922 rAlaAspGluValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSerProArgPr 942
Db 276140 -----CCACCAACAGCAGCTCACCGTCTACAGTCTCGTCAAGGGCAAGTGAAGGCC 276088
Qy 942 oSerLeuLeuThrProThr-----GlyAspProArgAlaAlaAsnAlaSerProGlnLysPr 960
Db 276087 GTCAAGAGCGGCACCTTACACGCGCGGTGACCCCGCGCACCCACACGACCCCGCGGCC 276028
Qy 960 oLeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaProProIleGlnValThrLy 980
Db 276027 G-----CGCGG 276022
Qy 980 sValHisGluProProArgGluAspAlaAlaPro-ThrLysProAlaProAlaProp 1000
Db 276021 CACCCACACAGCGCCCGCGCGCGCTCCACTCAGTCTGCTCACTTCCACACCCA 275962
Qy 1000 ro-----ProProGlnAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSerP 1019
Db 275961 CATGAGGCGCGCGGTGCACACCTTCGCGCAGCAGCTGGCCACGCGGTGT-TCCTCGG 275903

Qy	1019	roArg	-----GlyIysSerArgS	1025
Db	275902	CTCGATGTACGGGCTGATCGCCATCGGTTTACAGATGGTGTACGGCATCGTCCAGCTCAT	275843	
Qy	1025	erProAlaProProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyA	1045	
Db	275842	CAACTTCGCCACGGCGAGATCTTCATGACGGAGGCTTCGGCGCACTCAGGTCTACTT	275783	
Qy	1045	spProProCysTrpThrSerGlyLeuPheProValProArgGluValIleLysA	1065	
Db	275782	CTACGTCTCCAGACGGACAT--CCATGTGTAGTACCTCCCGGGATGCTCATCGG	275726	
Qy	1065	laSerProHisAlaPro-----AspProSerAlaPheSerTyrAlaProProGlyHisP	1083	
Db	275725	AGCGGCTCATCGCGTACTCATCGCGCTCGGGCGGAACGCTTCGCTACCGGC----	275670	
Qy	1083	roLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleS	1103	
Db	275669	-----CCCTCGGGGGCACACGCTCTCGCCCCC-----T	275639	
Qy	1103	erAenProProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleG	1123	
Db	275638	CATCACGCCCATCGGCTCTCCCTCGGCTCC-----	275607	
Qy	1123	lyAlaIleSerGlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLysA	1143	
Db	275606	-----AGCAGGTGTCTTCACTTCTACCTTCTACCGGACGCCA-----A	275573	
Qy	1143	laProValGlyProValThrMetGlyLeuProLeuPro-----	1155	
Db	275572	GACCGACAGACCTTCCCGAGCTGGGCACCGAGTCTTTCAACATCGGCTCGTCCACAT	275513	
Qy	1156	-----MetAspProLysLysLeuAlaProPheSerGlyV	1167	
Db	275512	CCAGAGGGCGCACTTTCATCATCATCGCGCCCCCATCTGTATGGCGCTCTCGGCTT	275453	
Qy	1167	alLysGlnGlnLeuSerProArgGlyAlaGlyProProGluSerLeuGlyValP	1187	
Db	275452	CTTCGTCCGCTCTCCGCGACGGCGCCATGACGGCACCGCCAGGACCGGACAC	275393	
Qy	1187	roThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyLys	1207	
Db	275392	CGCCACGCTCATGGGCATCGACA-----CCAACCGCATCATCGTCCGCTTCGCCAT	275339	
Qy	1207	erIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgG	1227	
Db	275338	CGCGGACTCTTCGCGCGGTGCGC-----CCGTGCGCTAGGCGCTCGGTACGGCCA	275285	
Qy	1227	lySerIleThr-HisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIle	1246	
Db	275284	GATCAAGTACGACATGGGCTTCCA-----	275261	
Qy	1247	IleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGly	1266	
Db	275260	-----GATGGGCTCAAGGCCTTCAACCGGCGCGT-----CCTCGCGG	275222	
Qy	1267	HisValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyMetSerVal	1286	
Db	275221	CATCG-----GCAACATCTACG	275205	
Qy	1287	ThrGlnCysSerLysGluAspGlyArgSerSer-----SerGlyProProHisGlu	1303	
Db	275204	CGCCATGCTCGCGGAGTGTCTCGCGTGCGCCGAAMACCCTCGCCACCGCTACATCG	275145	
Qy	1304	ThrAlaLapProLysArgThrTyrAspMetMetGluGlyArg-----	1317	
Db	275144	ACGGCATCCCGCGCATGCACGAGCTCGCGGGCGGCGAGCTGGGCCCTTCGGGCTTCG	275085	
Qy	1318	-----ValGlyArgAlaIleSerSerAlaSerIle-----	1327	
Db	275084	TCCTCTCATCTCTGCTGCTCGCGGCCCAAGGCGCTCTCTCGCGGAACGCTCGCGG	275025	
Qy	1328	-----GluGlyLeuMet	1331	

274038	Db	---	GCATGGAGGCGCTCACTCGTCTTGGCGGTCTCTCATCGCGCGCCCC	---GAGCTCCGCG	273985
1647	Qy	aAlaTyTyLeuProArgHisLeuAlaProAsn		---	1658
273984	Db	TGCGGCGGACTACCTCGCATCGTCACTTCGGCTTCGGTGAGATCTTCGGCATCGCG		---	273925
1659	Qy	-----ProThrTyProHisLeuTyProProTyLeuIleAr		---	1671
273924	Db	TCAACAGCGCTCGACGGCTCTCGGACCAACCTCACACGCGCCCAA		---	273874
1671	Qy	gGlyTyProAspThrAlaAla		LeuGluAsnAr	1682
273873	Db	GCATCTCCGGATCCCGACCTCGAATCTTCGATCAACCTCGGATCTCCACGACG		---	273814
1682	Qy	gGlnThrIle		IleAsnAspTyThrSerGlnGlnMethHisHis	1698
273813	Db	TGGCTCTCTCACCGTGGCGGCTTCGCACTACTTCTTCCTGATGCTGATCATCCCG		---	273754
1698	Qy	nThrAlaThrAla-MetAlaGlnArgAlaAspMetLeuArg		---	1714
273753	Db	CACCTCGTGTGTCTTCACAGGGCTCGGACTCCGGATCGGCGCTCTCGTGGATG		---	273694
1714	Qy	erProArgGluSerSerLeuAlaLeuAsnTyThrAlaGlyProArgGlyIle		---	1733
273693	Db	CCATCCGCGAGGAGACCGCGCCACC		GCATGGGCATCAACGGCTTCGGGTCA	273637
1733	Qy	spLeu-SerGlnValProHisLeuProValLeuValProProThrProGlyThrPro		---	1751
273636	Db	AACTCATCGCTTCGCCCTCG		---GCGCTCTCTCGCGCGCTCGCCGTA	273589
1752	Qy	-----AlaThrAlaMetAspArgLeu		---	1760
273588	Db	CGGTACGCGCCACGTCGGCTACAGCGTCAACCGCGCGCGTACAGTTCGCGGTCCG		---	273529
1761	Qy	LeuProThrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGly		---	1780
273528	Db	TGCCCCGGAATCCGCTTCCTGCTCGCGCGGTCTCTCGCGGATCGGCGACCGTCA		---	273469
1781	Qy	GlyProThrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArg		---	1800
273468	Db	AGGGCCCATCTCGCGCCACCTGCTACTCATCC		---CCGAGA	273424
1801	Qy	AspArgGluArgAspArgGluArgGluLysSerIleLeuThrSerThrThrThr		---	1820
273423	Db	AGCTCTCTTCCTGAAGAGTAGTACCAGTCTCGCTTCGGCATCGCTCTCTCTCA		---	273364
1821	Qy	ValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlySer		---	1840
273363	Db	TGGCTTCGCGCCCG		---AAGGCATCATCGCA	273334
1841	Qy	SerGlyGlyGlyGlySerSerSerArgProAla		---	1852
273333	Db	ACCGCGCGCGCAGCTCGAATTCACGAGACGACGATCGACGTACCGAACACGCC		---	273274
1853	Qy	-----SerHisSerHis		---	1856
273273	Db	TGACCGAACCACGTCGGGTCAAGAGCAGGCGGTGACCGACATGACACACACAC		---	273214
1857	Qy	AlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSer		---	1876
273213	Db	GGCCATCTCCACCGTCTCGAAGCCTCGCGCGTCAACATCGCTTCGGCGG		---	273163
1877	Qy	ValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGlnProSerLysProThr		---	1896
273162	Db	---ACTCACCGCGTACGCTCCGT		---CGACTCACGTCA-ACA	273125
1897	Qy	ValLeuArgSerThrThrSerSerProValArgProAla		---	1912
273124	Db	CGCGGAAATCGTCGGCTCATCGGCCCAACCGCGCGGCAAGACCACTTCTTCACT		---	273065
1913	Qy	PheProProAlaThrHisCysPro-LeuGly		---GlyThrLeuAspGlyValTyProThr	1931
273064	Db	GCCTCACCGGCTCTAGCTCCCAACGAGGAGGAGTCTCTCAAGGCGACGTCCTGC		---	273005

QY	1931	rLeuMetGluProValLeuLeuPro	-----LysG1 1941
DB	273004	CGCGAAGCCCA	-----CCTGGTGAAAGCGGGCATCGCCGACGTTCCAGA 272954
QY	1941	uAlaProArgValAlaAArgProGluArgProArgAlaAArgThrGlyHisAlaPheLeuAl 1961	
DB	272953	ACATCCGGCTCTTCGCCAAACATGACCGTCTCGA	----- 272920
QY	1961	AlaProProAlaAArgSerGlyLeuGluProAlaAArgSerProSerLysGlySerGluPr 1981	
DB	272919	-AAAGTCTCTAGGACGCCACACCAAGAACGCTCTGGTGGCCCTCTCTGC 272861	
QY	1981	ArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAs 2001	
DB	272860	GGCGCCCGGCTTCG	-----CAGGCCGAAACCGCTCCCGGAACGGCCATGAAC 272807
QY	2001	nLeuAlaProHisHisAlaSerProAsp	-----ProProAlaPro----- 2014
DB	272806	TCGTGAGTTATCGGCTCGCCACAAAGGCCGACACCTCTCCGCAACTCCCTTAAG 272747	
QY	2015	ProAlaSerAlaSerAspProHisArgGluLysThrGlnSerLysProPheSerI1 2033	
DB	272746	CGAACAGCGAAGCTCGAGATCGCCGCGCCTCGCCAGCAACCGGCTGCTGCTCC 272687	
QY	2033	eGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerTySerProGluGlyVa 2053	
DB	272686	TCGA	-----CGAGCGACGCGCGCAT----- 272665
QY	2053	lGluProValSerProValSerProSerLeuThrHisAspLysGlyLeu	----- 2070
DB	272664	-GAACCGCGAGAGACGCGAGC	-----GACCGAGAACTGGTCTTCG 272624
QY	2071	ProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArgProLy 2089	
DB	272623	CCATCCGACAT	-----GGGCATCGCGTCTCTCG 272594
QY	2089	sGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLe 2109	
DB	272593	TCAT	-----CGAGCAGCATCGCTTCATCTTCAACCTCT 272558
QY	2109	uProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysG1 2128	
DB	272557	GCACCGCTCGCGCTCT	-----CGTCCAGGCGGAGAAGC 272522
QY	2128	yHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyTh 2148	
DB	272521	TCATCGAAG	-----CACCTCGAGGTCTGTACAGGCGCAGCA-GC 272483
QY	2148	rArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTySerPheProG1 2168	
DB	272482	CGGTCTATCG	-----CGCCTTACCTCGCGCAACCTTCGAAGCGGCC- 272441
QY	2168	yAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyLeuPro	----- 2186
DB	272440	-CGGCGACCCGAGCGCCGAGGTCTGAAGCGCGCGAAGCCCAACCGGACCGCGGAG 272382	
QY	2187	ProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyG1 2203	
DB	272381	GCCGAAGCGCGGAAGCGGCGGACACACACCGAGACACGACCCACCGGACCGGACCGCAGC 272322	
QY	2203	yLysArgSerProGlu	-----ProAsnLysThrSerValLeuGlyGlyGlyLysAspG1 2221
DB	272321	ACCACCGGACCAAGGAGACGCCCACTGATGACCGCATCTGCTCGAGGTCTGAGGA- 272270	
QY	2221	yIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVa 2241	
DB	272269	-----CCTCAGGCTCTCTTACGCGCAAGATCGAAGCGCGTCAAGGG- 272231	
QY	2241	lTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMetGlySerLysS 2261	
DB	272230	-CATCTCTTCAGCGTCGAAGCGGCGC-AGGTGCTCACCTCA-----TCGGCACCC 272182	


```
Qy 85 -----ArgProGluSerHisSerTyrLeuProGluLeuGlyLysSer 98
Db 25936 GCAGGGGCGCCGGGACGCCACACGGCTGCAGACATGATGCCAACAGCAACGGCGGC 25995
Qy 99 GluMetGluPheIleGluSer-----LysArgPro-----ArgLeuGluLeu 112
Db 25996 CGGTATCGCCATTCTCGCGGTGTAACGCAAGCGCCCGGTAGGTTCGTGATTCGCA--- 26052
Qy 113 LeuProAspProLeuLeuAtrqProSerProLeuLeuAlaThrGlyGlnProAlaGlySer 132
Db 26053 -----GCTCGCCCTGTGTGAGCCCGCGTTCATGCGCGCCACCGAGGGGG 26097
Qy 133 GluAspLeuThrLysAspArgSerLeuThrGly-----Lys 144
Db 26098 CGTCAAGCCCGCGGCAACGGCGCGTAGTGTGCGGTGTCGGCGCAGCAGCGCATCGG 26157
Qy 145 LeuGluProValSerProProSerProProHisThrAsp---ProGluLeuGluLeuVal 163
Db 26158 CATCGCGGATTCACGCGAGTTCTTCCAAACGATACCTGGCGCACCGTCTCTAGCAGGCGGT 26217
Qy 164 ProPro-----ArgLeuSerLysGluLeuIleGlnAsnMetAspArg 178
Db 26218 AGCGCGTTCGCGCGTGTGTCGTGCGTACCCACAGAGACTTGTCTGCCAGCAGGCTGA 26277
Qy 179 ValAspArgGluIleThrMetValGluGlnGlnIleSerLysLysLysLysGlnGln 198
Db 26278 GCAGATCGACCATCTCGTAGCGCTGAACGTCACCGCGCGGCTGCGCTTGGGCACCGT 26337
Qy 199 GlnLeuGluGluAlaLysProProGluProGluLysProValSerProProPro 218
Db 26338 CGAGATCAAAACCCGCTCGGGAACCCGCACTGCCGAA-----ACAGCACTGTCTCG 26391
Qy 219 IleGluSerLysHisArgSerLeuValGlnIleIle-TyrAspLysAsnArgLysLysAl 238
Db 26392 GTC-----CGGTACAGCAGCGCATGTGACCATCGAGCGAAGCCCGCATCGTCT 26439
Qy 238 aGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAs 258
Db 26440 GCTGCGCGGCGCACGCAATACCGCATCCACCGTCCAGCGCGGAA---CGGTATCGCA 26496
Qy 258 nGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLy 278
Db 26497 AGCTGTCCAGCATTTCCGTGAG----- 26518
Qy 278 sLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPh 298
Db 26519 -----CGCCAGGCGACGCAACCCCGCAGC----- 26542
Qy 298 eCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGlu----- 314
Db 26543 -TGC---AAGTTCGATCGCAGCGGATGCGGTGCGTCCAGTCCGTGGCAGATCTCGGTCA 26598
Qy 315 -----ArgIleGluAsnAsnProArg----- 321
Db 26599 GGGCAGAGTTGTCCGAGTGTATCTCGAGTTCCGGCCCGCGCTCACAGCGCGGTGCGTGA 26658
Qy 322 -----ArgArgAlaLysGluSerLysValArgGluTyrTy 333
Db 26659 ACAACTCGATCGCTCGCCGTGCCAGCGGGGAACCCGCCAAATCTGTCTACCCGCCA 26718
Qy 333 rGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSerAr 353
Db 26719 CCGGATCGGTTCCCGGCTGTGCGCAATACCTTCAGCGTGGCGACGCCCGGAGCAACG 26778
Qy 353 gValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHis---GluVa 372
Db 26779 CGAGCATCAGCGCGCGCCGTCGAGCAAGTCTCGCAGTGTCTCCAGCACTACCAGCA 26838
Qy 372 lSerGluIleIleAspGlyLeuSerGluGlnGlnAsnLeuGluLysGlnMetArgGlnLe 392
Db 26839 TGC CGCGCTGCCGAT-----ACGCCGCAAAATGGTGTCCACGTCGACGCGCGCGCT 26892
```

```
Qy 392 uAlaValIlePro-----ProMetLeuTyrAspAlaAspGlnGlnArgIle----- 407
Db 26893 GATCCGGCAACCCCAAAACCCGCGCGCGATCGGCACCGAGCCCGCGGTGCGTGCATCG 26952
Qy 408 -----LysPheIle 410
Db 26953 GCGCCAGTTGACATACCAAAACCCCGTCGGATAACCGTCGGCAACCGCGCTCCGACCT 27012
Qy 410 eAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAs 430
Db 27013 GTGTGCCAGCGGTCTTTCCAGACCCCGCGGACACCGGTAAAGGTGACCCACCGTTGA 27072
Qy 430 nMetTrpSerGluGlnGlnLysGluThrPheArgGluLysPheMetGlnHisProLysAs 450
Db 27073 CGTC-----CAGCAGCCACGAGCTTGGCCACTTGTCTGACGCGGCCACCA 27120
Qy 450 nPheGlyLeuIleAlaSerPheLeuGluArgLysThrVal----- 463
Db 27121 G-----CCGAGTGAGCTGGCGCGGAGACAGACAGTGCACCAACGACTTTCGGG 27168
Qy 464 -----AlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLy 480
Db 27169 TCCGACGCGCGGAAACGCGTGTGTGAGATCAGGTCACAGCTGCACCCCGTTCG 27228
Qy 480 sSer-----LeuValArgArgSerTyrArgArgArgGlyLysSerGl 494
Db 27229 GTCGGGCAGTCTGTCAGCGGTAGGTACCGAGGTGTTTCAGCCACGCTCTTGGGCA 27288
Qy 494 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 514
Db 27289 GCAGGTGAGCAACACAGATCTGCTGTAGTTCCTCCGACACACAGCTCTGCGCCCGG 27348
Qy 514 gSerSerGln----- 517
Db 27349 GTCGCGCACCGCGGTCGCGTGTGTCGCGCTTCAGCAGTTCCTCTCGCGGTG 27408
Qy 518 -GluGluLysAspGluLysGluLysGluLysGluAlaGluLysGluGluLysPro-- 536
Db 27409 ACGACACCTCCCGGTGTGTCATCCGATGCGCAGCGGATCGGTGCCAGCGGCCGCT 27468
Qy 537 -GluValGluAsnAspLysGluAspLeuLeuLysGluLysThrAspThrSerGlyGl 556
Db 27469 GCAAGCCCGAGGCGCGCACCGCCCGCTC-----GGATCGCGCGCGCA 27510
Qy 556 uAspAsnAspGluLysGluAlaValAlaSerLysGlyArgLysThr-----Al 572
Db 27511 ACCCCACCAAGAGTGTGCTGCTTCGCCCTGTTCGACCGCGGCAACCCCGCGGTGTCGC 27570
Qy 572 aAsnSerGlnGlyArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSe 592
Db 27571 GAACCAATTCCGTGAGGTTC-GGTCCAGTTTCGCGATCCCGTCGTCGTCGTCGTCGAG 27629
Qy 592 r-----GluGluAlaIleThrProGlnGlnSerAl 602
Db 27630 CCGCGCAGTGGTTCGCGCTCGATATCGCGCCAGCAGCAACGTCACCGTCGCGCTCG 27689
Qy 602 aGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGlu----- 618
Db 27690 ACAAGCTCGTCAACCATCTGCGTCCAGTCCACAGGT-ACCACGTGACGCGCGGGTG 27748
Qy 619 -----GluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyAr 632
Db 27749 AATCTTCTCATGTACCGACGATCGAGCCAGCGGTAGCGCATTCATCGGCAC----- 27803
Qy 632 gAsnTrpSerAlaIleAlaArgMetValGlySerLysThrVal-----SerGlnCy 649
Db 27804 -----CTGCGCTAGATGCTCGAAATCTTTGGCGCGCGCTCCATGTG 27847
Qy 649 sLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGlu----- 664
Db 27848 TTCTACGCGCTTTAGTCGATGCAATTCGGCCAGCGCGCTGCCATCGCGCGGAGCCTACAGT 27907
Qy 665 ----IleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArgArgLysLy 683
```


Db	27908	GCCTGCTCTGCTGGCAATTTGAGTCCCCAGTGTGGCAGCATCCCGCGCAAGAA	27967	Db	28876	CGTTTCCGCGCGCGCCACCGCCGCTAGCACCAGTGTGCGCGCGCTGCGCGCGTGC	28935
Qy	683	sLys-----LysAlaProAlaAlaAlaSerGluGluA	694	Qy	1005	-----LeuGlnProGluSerAspAla	1011
Db	27968	CCGACACGCGCATCGTGGCGGTGCGTGCAGAGTCGCGCTACCGATCGCGGAGT	28027	Db	28936	CGCCCGCTGTTCTTAATGTCAAAGCCGTCAACGCCGTTACCGCGGACCCACAGCC	28995
Qy	694	la-----AlaPheProProValValGluAseGluGluMetGluAlaSerGlyV	710	Qy	1012	ProGlnGlnProGlySerSerProArgGlyLysSerArg-----SerProAlaPro	1028
Db	28028	CACCGGCCGATGGGTTTCCGAAGCTAGGATGATGATGATGGGCGCGCCCTCGG	28087	Db	28996	CGCCTTAGCTCCGCGCCCGCGCTCACACCGTGTGAGTCTCGCGCGCGCCA	29055
Qy	710	al-----SerGlyAenGluGluGluMetValGluGluA	722	Qy	1029	ProAla-----AspLysGluAlaPhe	1035
Db	28088	CGCCTTTCGCGTAACCCCAACCATCGCAAAACGAGCCTAGCTGCCCGCGCGCA	28147	Db	29056	CGCGCCCGCGCTGCTGCTCCGCCACCGCTGCCCGCTCGAAGCCGCTTT	29115
Qy	722	luAlaLeuHisAlaSerGly-----AsnGluValProArg-----GlyG	735	Qy	1036	AlaAlaGluAlaGlnLysLeuProGlyAspPro-----ProCysTyrThrSer	1051
Db	28148	GAGGAGCCATTCGCGTGGCGCCCAACACGACGAGGATCCGATCTCTGACCATGGT	28207	Db	29116	AGTCTCGAGTCTGGGCGCACCAACCCGCGCCACCGCGCGCGTGTCTCTCTGGCGG	29172
Qy	735	luCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerProH	755	Qy	1052	GlyLeuPheProValProProArgGluValIleLysAlaSerProHisAlaProAsp	1071
Db	28208	GGTGTCTGCTAGCTACGTAA-----ACGGAGATGGAGGCGCCTTCGCGG	28257	Db	29173	-----CCGCTTCGCGCGCGCTTCATCCCGCGTGTGCGCGCGCGCGCGCG	29220
Qy	755	isThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyA	775	Qy	1072	ProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThr	1091
Db	28258	-----CCTTCACCGCGG-----	28269	Db	29221	-----TGTCCACCAACCCCGCGCGCGCGCGCGCGGCTG-----	29253
Qy	775	laAspGlyProProGlyProProThrProProArgArgThrSerArgAlaProIleG	795	Qy	1092	AlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSerSe	1111
Db	28270	--GAACCGCGGAGCCAGGTCGCTCCCTGTCGCGGAGCACCGACGATCCCGGACG	28327	Db	29254	-----CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC	29295
Qy	795	luProThrProAlaSerGluAlaThrGlyAlaProThrPro-----ProProAlaProp	813	Qy	1111	rAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerVa	1131
Db	28328	AGCGAGCGCGCGGAGTCAAGCCACCGCACCGCTCGCGCTCACCTTCGCGCGCGC	28387	Db	29296	CGCGCTGCGCCCAAGAAATGCTCGCGCGCGC-----	29326
Qy	813	roSerProSerAlaProProProValValProLysGluGluLysGluGluThrAlaA	833	Qy	1131	lGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGl	1151
Db	28388	CGTCCCGCGCTCACCGCGCGATGTGCG-AGCGGAGGGCGCGCAAGCGCGCGAG	28446	Db	29327	-----CACCAAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG	29378
Qy	833	laAlaProPro-----ValGluGluGlyGluGluGlnLysProProAlaAlaG	849	Qy	1151	yLeuPro---LeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGl	1170
Db	28447	CGCGCGTTCGCGCGTTCGCTCACACCGCGGCGCGCACCGCTTCGCGCGCGCGCCA	28506	Db	29379	CCCTCGAGGGCACCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCA	29424
Qy	849	luGluLeuAlaValAspThrGlyLysAlaGluGluProValLysSerGluCysThrGluG	869	Qy	1170	uGlnLeuSerProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGl	1190
Db	28507	CCCAAGCGCGGT---GCCGCGCTTCCACCGCGCGCGCGCGCGCTGCTGATCAGCAA	28563	Db	29425	-CGCGCTCACCGATGCTCGCGCGCGTCAAGCGCGC-----TTGCTCCAGCGCC	29471
Qy	869	luAlaGluGluGlyProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAlaGluG	889	Qy	1190	nGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLy	1210
Db	28564	CTGCGGTGCGCTCCCAACCGTGACCGCGTTCGCGCGCTCGCTCACCGCGCGGG	28623	Db	29472	ACCGGCGCCG-CGCATCACCGCGAGAGCTTCGCGCGCGGTGCGCGCGCGCGCGCGCG	29530
Qy	889	lyAlaLeuLysAlaGluLysGluGlyGlySerGlyArgAlaThr-----T	905	Qy	1210	sGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleTh	1230
Db	28624	GCGCC-----GAAGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	28668	Db	29531	CGGCTCGCGGTCCCGAATAGCCCGCGGTTCGCGCGGTTCACCGCGCGCGCGCGCGCG	29590
Qy	905	hrAlaLysSerSerGlyAlaProGlnAspSerSerSerAlaThrCysSerAlaAspG	925	Qy	1230	rHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleGlyGluAs	1250
Db	28669	CGGCTGTTCGCGCGGCG	28728	Db	29591	CGCGCGCGCGCGCGCG-----CTTGGCAGCGGTGCG-----	29621
Qy	925	luValAspGluAlaGluGlyGlyAspLysAsnArgLeuLeuSer-ProArgProSerLeu	944	Qy	1250	pSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTy	1270
Db	28729	GGCTC---TCCGTCAAAATCCGAGATAGTAAAGCGTCAACCGCGCAAGCACCATTA	28785	Db	29622	-GCCGACGCGCGTCCCGCGCGCGTTCAGTGGCGCGCGGTTCGTGGCGGTTCACCGCGT	29678
Qy	945	LeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLys	964	Qy	1270	rGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSe	1290
Db	28786	CCAGCGTCCCG	28836	Db	29679	----GACCGGAGGTTCGCGGTTCGCGGTTCACCGCGCGCGCGCGCGCGCGCGCGCG	29734
Qy	965	GlnLeuLysGlnArgAlaAlaIleProProIleGlnValThrLysValHisGluPro	984	Qy	1290	rLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgTh	1310
Db	28837	TTGCCGACCTCGCGCGGTGCGACCGCGG-----GCCCGC	28875	Db	29735	C-----GGGGTCAACCGCGGTTCACCGCGCGCTC	29761
Qy	985	ProArgGluAspAlaAlaProThrLysProAlaProProAlaProProProGlnAsn	1004	Qy	1310	rTyrAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLe	1330
Db				Db	29762	CGCGCGCGCGCGCGCGGTGATACCAATGCGCGCGT---GCCGCGCGCGCGCGCGCGCG	29818

D	B	31368	GGCACCA	CGTCCGCGCCA-----ACACCTGG-----	31395
	Q	Y	2035	uLeuGluLeuArgSerLeuGlyTyxHisGlySerSerTyxSerProGluGlyValGluPr	2055
	D	B	31396	-CTGCCCGCGCGA-----CCCGAGGTGGTCACC	31424
	Q	Y	2055	oValSerProValSerSerProSerLeuThrHisAspLysGlyLeuProLysHisLeuGl	2075
	D	B	31425	GCACGCCCAACCGGCCCAACGTCACCGCGCGCACCGCGGCTCGCATCA-----CC	31478
	Q	Y	2075	uGluLeuAspLysSerHisLeuGluGlycyluLeuArgProLysGlnProGlyProValLy	2095
	D	B	31479	ACGGTTGCCCGCGTCAGATACGAGCACAGCATTGAACCGTGAGTCGGTTACCA----	31533
	Q	Y	2095	sLeuGlyGlyGluAlaAlaHisLeuproHisLeuArgProLeuproGluSerGlnProSe	2115
	D	B	31534	-----CGGGCCCCCGCGCCCCACCGTTGCCGCGCGCACCGCGCGCGCCGCGCATC	31583
	Q	Y	2115	r-----SerSerProLeuLeuGlnThrAl	2123
	D	B	31584	GCGGGCGTGGCCCCCAACCGCGCGCGCGCCCCACCGGCCCGCGGTACTCTCCATCCTC	31643
	Q	Y	2123	aProGlyValLysGlyHisGlnArgValValThrLeu-----AlaGlnHisIleSerGl	2141
	D	B	31644	ACCGGGGTACCGGATGACACCCAGGAAGATCCCGCTCATATCGGCATAGCGGCACCGC	31703
	Q	Y	2141	uValIleThrGlnAspTyxThrArgHisHisProGlnGlnLeuSerAlaProLeupro--	2160
	D	B	31704	GGTCCGCTACCCGCCCATGACCGCGCGCCCCACCGTCACCGAACAAACCGCGCGTTGCCAC	31763
	Q	Y	2161	---AlaProLeuTyxSerPheProGlyAlaSer-----CysProValLe	2174
	D	B	31764	GGCACCGCGCGTTACCGCATGACCGCGCGGACCGGTGGTGGCGCGCATTCGCCACTTT	31823
	Q	Y	2174	uAspLeuArgProProSerAspLeuTyxLeuPro- ProProAspHisGlyAlaProA	2194
	D	B	31824	GCGCGCGTCCCGCGCTGACCGCGCGCGTGGTGGCGCGCGCACACCATTTG-GCCCGCG	31882
	Q	Y	2194	laArgGlySerProHisSerGluGlyGlyArgSerProGluProAsnLysThrserv	2214
	D	B	31883	CAGCAGACCG	31935
	Q	Y	2214	alleuGlyGlyGlyGluAspGlyileGluProValSerProGluGlyMetThrGluP	2234
	D	B	31936	-----CCGGGTGCGCACCGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	31966
	Q	Y	2234	roGlyHisSerArgSerAlaValTyxProLeuLeuTyxArgAspGlyGluGlnThrGluP	2254
	D	B	31967	CGTCACCGCGCGCGCACCACTGCCACCGTTG-----CCGCGCGCACCGCGCGCGCG	32011
	Q	Y	2254	roSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerL	2274
	D	B	32012	CGTCACCGCGCGTCCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	32071
	Q	Y	2274	ysLeuThrGluSerAsnSerAla-MetValLysSerLysLysGlnGluIleAsnLysLys	2293
	D	B	32072	CGACATACCCCAATAGTCAGCGACCCCGCGCACCCCGGGCGCGCGCGCGCGCGCGGT	32131
	Q	Y	2294	LeuAsnThrHis-----AsnArgAsnGluProGluTyxAsn	2305
	D	B	32132	TTCGCGCGGCACCGCCATTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGATACCG	32191
	Q	Y	2306	IleSerGlnProGlyThrGluIlePhe-AnmMetProAlaIleThrGlyThrGlyLeuMe	2325
	D	B	32192	-----CGGGCGCAATTTATTTCGGGTAGCCCCCGCGCACCGCGCGCGCGCGCGCG	32239
	Q	Y	2325	tThr-TyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIl	2345
	D	B	32240	CACCGGCACACCG	32294
	Q	Y	2345	eIleArgLysAlaLeuMetGlyLysTyxAspGlnTrpGluGluSerProProLeuSerAl	2365

Db	32295	-----GAGAAGCCACCTCCCGGACCAC	32311
Qy	2365	aAsnAlaPheAsnProLeu---AsnAlaSerAlaSerLeuProAlaAlaMetProIleTh	2384
Db	32318	CGTTTCGGCCGCCCCACCGGACCGCATCGCCCGCGACCTGCCAAACCGTCAC	32377
Qy	2384	rAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLy	2404
Db	32378	CGCGCGC-----ACCGTCGTCGACCCCGGCACACACCGCCGCCCAT	32419
Qy	2404	sValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerG1	2424
Db	32420	CCTGACCGGCGGTAGCAGCTGTGGCCCGCGCGACCTACACCCACCGCGGACGCGC	32479
Qy	2424	y-----AspArgProSerValSerSerValHisSerGluGlyAspCysAsnArgAr	2442
Db	32480	CCCCATGACCAACACCCCGCGGTATCCGCC---CGCGCGCGCGCGCACACACCGAGCC	32536
Qy	2442	gThrProLeuThrAsnArgVal-TyrGluAspArgProSerSerAlaGlySerThrProP	2462
Db	32537	CACCGGCACACCGGCGCGCGCTTGTCCGAACAGGCGCGGTGCCACCCCGCGCGCGG	32596
Qy	2462	hePro-----TyrAsnProLeuI	2468
Db	32597	CGCCACACACGAGCCGCCACCCATACCGCGCGCGCCGCCACCGCAGCATCCACCGG	32656
Qy	2468	leMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlys	2488
Db	32657	TCCCGCGGTACCTCGGGTGGCGCCCGCGCGCGCACCGCGCGCGATCCCAA	32716
Qy	2488	erGlyProLeuAlaGlyPro	2494
Db	32717	ATRACCGCGCGCGCGCGC	32736
RESULT 91			
AV211023/c			
LOCUS	96742 bp	DNA	linear BCT 03-MAR-2003
DEFINITION	Streptomyces violaceoruber strain SANK95570 plasmid pSV2, complete sequence.		
ACCESSION	AY211023		AY055228
VERSION	AY211023.1		GI:28797237
KEYWORDS	Streptomyces violaceoruber		
SOURCE	Streptomyces violaceoruber		
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
REFERENCE	1 (bases 1 to 96742)		
AUTHORS	Spatz,K., Kohn,H. and Redenbach,M.		
TITLE	Characterization of the Streptomyces violaceoruber SANK95570 plasmids pSV1 and pSV2		
JOURNAL	FEMS Microbiol. Lett. 213 (1), 87-92 (2002)		
MEDLINE	22123362		
PubMed	12127493		
REFERENCE	2 (bases 1 to 96742)		
AUTHORS	Spatz,K., Scholz,C.J. and Redenbach,M.		
TITLE	Complete nucleotide sequence of the linear plasmid pSV2 from Streptomyces violaceoruber SANK95570		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 96742)		
AUTHORS	Spatz,K. and Redenbach,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-SEP-2001) Department of Genetics, Genome Research Unit, Kaiserslautern University, Erwin-Schroedinger-Str., Kaiserslautern 67663, Germany		
REFERENCE	4 (bases 1 to 96742)		
AUTHORS	Spatz,K., Scholz,C.J. and Redenbach,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-DEC-2002) Genome Research Unit, University of Kaiserslautern, Erwin Schroedinger Str. 24, Kaiserslautern, Rheinland-Pfalz 67663, Germany		
REMARK	Nucleotide sequence updated by submitter		
COMMENT	On or before Mar 3, 2003 this sequence version replaced gi:22096410, gi:22096411, gi:22096409.		

FEATURES
sourceLocation/Qualifiers
1. .96742

/organism="Streptomyces violaceoruber"

/mol_type="genomic DNA"

/strain="SANK95570"

/db_xref="taxon:1935"

/plasmid="psv2"

655. .1137

/gene="psv2.01"

655. .1137

/gene="psv2.01"

/note="similar to hypothetical protein from Streptomyces coelicolor A3(2) encoded by GenBank Accession Number AL939132"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="hypothetical protein"

/protein_id="AA050085.1"

/db_xref="GI:28797238"

/translation="MELDGMAPRSRGVFGDWMDEGVTTVSEALVGIADRDWRRGGR SAGAAELAVHADVOKREGAGLVLEFVRGHVLPRTHTPTTARHVRVLEALGTGGLC TARTNADGVVWCTLDAGHYDADDKPPKDPKPGGWHKADASTWINDIAGACIPHAAL"

gene

CDS

1233. .3836

/gene="psv2.02"

1233. .3836

/gene="psv2.02"

/note="similar to Streptomyces lividans helicase-like protein and Streptomyces coelicolor A3(2) putative helicase encoded by GenBank Accession Numbers AF194023 and AL590463, respectively; DEAD-like helicase superfamily; helicase associated domain"

/codon_start=1

/transl_table=11

/product="putative helicase-like protein"

/protein_id="AA050086.1"

/db_xref="GI:28797239"

/translation="MGPPARSVPVQPGGARSQRLARARRSPACALSFVDRGIIV TVPTDLAQTAAQAVRAGHRS PMVAVCSLENDPMLNOLGVRTTNPILQALWAGHGP VVVFATYLSAVLEDVDAPSGQKVRGPLEALAGGERLYGQMDGFDLAIIDEANGT AGDLGRPMALHONARI PADFLYLTATPRILAAAPKQAGQAEIATMADDPNGT YGAWLAEGLSEALIERILAGFEIDVLEIRDPSPVVGTEEARRRRLALLQALLLEH AAYSLRTVNTFFQKVEEAAAFADKLPETAELLYVNDASDADLAAADKLPASSIDAEF YELAGRHVPDPVRSAWLCGDHVAERREKIROFANGINAAGRVHRALFASVRVLG EGVDI TCGERGVEAVSPADTRGSQVEIVQNI GRALRLNRDGS TKVARIIVPVELEPNE D PAMVASASPKPLVAVLQGRSHDERLVEQLASRALTSGRKRVHVDDEGRIVGAGG ESDEGDEQEQGDQDDTDAAEASALLHFSFPRVATIAAFIRTVRTPESLIVWLEGT ALLRWABNGITGVHAPYDTEVEVGVTDFPIGRWVHQQRKALRAGELEDRRKTLLD APAGWMEPEGEAWENKLAALRSYRRMGHLAPRODAVWGEGEAMVPVQGHMANLR KGGIGKDAERAARAQOLAAVDEDWNCWPLDWQRVYRVILADVLADGVLPALPEPGLV FEGDDICRWLEROKNPGNWAQLSTDQOERLTILGVTVPVAPTAKGASKGQGAQAFQ RGLAALAOYIARBEGHVRPAHAAEIVAEGEAASVVVVLGVWVWNTKSRDKLTATQOL AALAKLGVDA"

gene

CDS

3925. .4764

/gene="psv2.03"

3925. .4764

/gene="psv2.03"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="hypothetical protein"

/protein_id="AA050087.1"

/db_xref="GI:28797240"

/translation="MAGTTEPAPLWAEGVDPDHLAVPVROWLYGVLRDYSLAARVAVWL KLFSLHLDTPQATLAATFAEDETNPMRLSIIIDATIGCLRHVLETAHHSIEIGAAASV MELLEILHEDGSAFTSRDGSGLWRINETLHATYKAVEAGASMAQTAADHLRAAPS EAYGIKDPDGAAYSRAIKAVEAVSPLFNAPETPTIGKVRSHLDQGRHXYMVIADK TGA PASIDA VAVMISLIMHGOQRDRHEGGPT SAPVTQEAATAVHTAAILIHWISNGSI QKK"

gene

CDS

4815. .5306

/gene="psv2.04"

4815. .5306

/gene="psv2.04"

/note="similar to hypothetical protein from Streptomyces coelicolor A3(2) encoded by GenBank Accession Number AL939104"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="hypothetical protein"

/protein_id="AA050088.1"

/db_xref="GI:28797241"

/translation="MORGCPOMPQGVPCCLATTSEGHFSQVGSASFGLSPGWAGLTK ALQWVEREGARPVFRGHTTEETVYVSEAFVWLKGVWISNTKPGGAGLPRNWTWTC GARHGLMGIMGHYGAAGAPRRPARGQTWHPARAASAGMIRAVRTIDERSYIGALLCV CRR"

5429. .6127

/gene="psv2.05"

5429. .6127

/gene="psv2.05"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="hypothetical protein"

/protein_id="AA050089.1"

/db_xref="GI:28797242"

/translation="MRIRASRWQDAVRMLLLNAAAHPRPQVDGVPACVGVVRA QVRLQKLDPMVRYPDYLAELMTEYKAPDEPALLDLAEDILSSQEPDLRRFPMRLHR QGAFELDNALAVLVERGLLKTQVLRKRVKVEHRYLLQAGRTLVNI VVEAPALAW IVERTKRVVALVDGLGTLKKRQYLLRDVADTPIGSYITPTIDQARARLARALRAIAA PQASAPADQEQAS"

6124. .8130

/gene="psv2.06"

6124. .8130

/gene="psv2.06"

/note="similar to putative large ATP-binding protein from Streptomyces coelicolor A3(2) encoded by GenBank Accession Number AL939116"

/codon_start=1

/transl_table=11

/product="putative large ATP-binding protein"

/protein_id="AA050090.1"

/db_xref="GI:28797243"

/translation="MTKPDVPATVPVPTADDILTQAARLDVAKHVERALAEANITA LTPPAQRRRLRLRLRVGIDHGEFTTIDRPFPGIWAIVHPENSAGKTSLLLEFLV WPLRGAPDLPDGVRSWVRLHLDDVVAGPRTIALEQIPGLLPVPSVDILTADTVQA LLAADDNPLRLVAKGSGTDDVERVIGSPFLDALRMERTSLWATGADGEGAPQVHG WPAVFGACVLYNACGDELLIGDVSAPGLPAKLERFVDIPIYSSAHQOLSVAVKRAKTT RQQRRAEGDLHARSARQAQQLDLMOQEIQLSGAPETDIAPLQRAQQAQELH RLMSLNGAQSLSADAAAARQAGQDQAGTQWARRVLRNLNPSCCPCEEPVGE RRAERDRATCAVCTRPLPEIAEDIAEAVLEQLTTDIATALEAEAAAHHRAVQASA DAATVAHQQAATEAVSALAASPAHVRLELELTAELKGLRLAATGTPATPTVSAVP DVLVAHQVLEAVTKASADYLPFTMDSEIVLSQFCGVQHLNSTVTKRNGNVNPKEG VKHPFKLSRGDRLMRVATVIALLRAGAARHIDSHPLGLLLIDSVAAAEEMTADAARLL IDELQTLTPELPELVQVFTTAQPELVEGLLPEDHIIITSDSEHLF"

8299. .10257

/gene="psv2.07"

8299. .10257

/gene="psv2.07"

/note="similar to hypothetical protein from Streptomyces coelicolor A3(2) encoded by GenBank Accession Number AL939116"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="hypothetical protein"

/protein_id="AA050091.1"

/db_xref="GI:28797244"

/translation="MAAGPOLPYLLTALAHGIDITI PAEQBARLLAAATQGLRQPHSAW VLADALDVCTREFGLAHLGNRIVRDLTTLAEADLTGCDAAQAQAIAGLLRLAVAG GSPHRLALLTETITGETGPADALDRPLIGIARDHFADTDRLLDNLNALSQDLNPA TRADASFELALADERALRAADRTTAGEOLQORALRMFTELDRTHEARLARAHAATE AVLAFADLEONFTGDS SARQLTEAADQLDAAHLAAWTGRMHOLDWLSARSLTQSA WAKLVTLNTAQAHLDQPSWNPADLNRLLEIYLTSSR IHAPAKOSTGLUTLSIPAV EAPFVRNHGILLHQLEALHPDQPTDHPDAQTLYDAVVRQRRLQSSQAQOQPEAMPGVK

QY 1004 AsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer-ProArgGlyLysse 1023
Db 51102 AAC - - - - - CCCCCTCTTCGGCCCGCGTGCAGTG 51073
QY 1023 rArgSerProAlaProProAlaAspLysGluAlaPheAlaGluAlaGln-LysLeuP 1043
Db 51072 CCCTTCGGCCCGCCCG - - - - - CAGACGCTCC 51046
QY 1043 roGlyAspProPro - - - - - CysTTPThrSerGlyL 1053
Db 51045 CCGCCACCCCGAGAGCTTTCACCCGCGACGCGACCGACCCGATGACGTATCGC 50986
QY 1053 euProPheProValProProArgGluValIleLysAlaSerProHisAlaProAspProS 1073
Db 50985 GGCCCTGCGCGAGCCG - - - - - GTCTCGGCACCCCTTCACCGCGCAATCCGA 50938
QY 1073 erAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaA 1093
Db 50937 GC - - - - - CCGC 50932
QY 1093 rgProValLeuProAtgProProThr - - - - - 1101
Db 50931 GCGCGCGCCCGCGAGCCCGCGAGCGGGCGGTCTGGTCTCCGGCGACCTTCGCGTC 50872
QY 1102 - - - - - IleSerAsnProPro - - - - - ProLeuIleSerSerAlaL 1113
Db 50871 GGGGGTTCGACGCCCGCGCGCGCTGGTTCCTCTCGCGCGATCTCTCGC 50812
QY 1113 yHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnL 1133
Db 50811 GGTCTCTCTCGCACTA - - - - - TGC CGG 50788
QY 1133 euHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuP 1153
Db 50787 CATTCCTTCGCGTTCGCGAGACC - - - - - GGCCACACCGGAGCAGCGCGAAGCGGTTCG 50731
QY 1153 roLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuS 1173
Db 50730 GAAGACCGCGGCACCCCAAGAGCGGCCCGCGGGCGGTTCGAAGAAGCAGGTGCA 50671
QY 1173 erPro - - - - - ArgGlyGlnAlaGlyProProGluSerLeuGlyV 1186
Db 50670 GCCCAGCACGAGTACGCGCGCAGACGCTT-CAGCGCGTCTCT - - - - - GGCG 50624
QY 1186 alProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyG 1206
Db 50623 GGCG - - - - - GGAGCCCTCTCCGCGCGTCTCCGGTCCGGCGCA 50588
QY 1206 lySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyrA 1226
Db 50587 GGTACGGCGGGCGCGCGCTCTGTAC - - - - - AGCGGTTCGGAGCGCGGGG 50537
QY 1226 rgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgI 1246
Db 50536 CGCGCGGTTCAGCACCCGCTCGCCAGCGC - - - - - TCCGG - - - - - 50499
QY 1246 leIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysG 1266
Db 50498 - - - - - GGCGCGCGCGCGGTTCAGCGGGCGCGCGCGGTTCGGTCTCCGGCGCT - - - - - G 50447
QY 1266 lyHisValIleTyrGluGlyLysGlyHisValLeuSerTyrGluGlyGlyMetSerV 1286
Db 50446 GCGCTCCAACTCCGGCGGGGCAAG - - - - - GGCGGTTCGGG 50408
QY 1286 alThrGlnCysSerLysGluAspGlyArgSerSerSerGlyProProHisGluThrAlaA 1306
Db 50407 TCACCCGCTCGGC - - - - - GCACCTCGCAGCGCGCGAGCTGAACCTCAAGAAGAAGCCGCG 50351
QY 1306 laProLysArgThrTyrAspMetMetGlu-GlyArgValGlyArgAlaIleSerSerAla 1325
Db 50350 CCCCGATCTGTCG - - - - - GCGGCTCGCGCGGTTCGGGAGGTTCGCGCGCGGCA 50300
QY 1326 SerIleGluGlyLeuMetGlyArgAlaIleProProGlu - - - - - 1338

Db 50299 GGGCCCGAGCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 50240
QY 1339 - - - - - ArgHisSerProHisLysLeuLysGluGlnHis 1349
Db 50239 CGGGTCCGGCTCCCGCGGGGAGACGACGACGACGACGACGACGACGACGACG 50180
QY 1350 HisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAsp 1369
Db 50179 CGGGCGCGGGG - - - - - CGGCG 50162
QY 1370 TyrLeuArgGluAlaLysLeuLys - - - - - ArgGluGly 1382
Db 50161 TTCTCTCGCGGTTCGGCGAGGAAACGCGACCGCGCGCGCGCGCGCGCGCGCG 50102
QY 1383 ThrProProProPro - - - - - ProSerArgAspLeuThrGluAlaTyrLysThrGln 1400
Db 50101 GA-CGCGCCCGGCACCAAGACCCCGCAAGGCGACGGGGTTCGGGAGGAAGCTTCG 50043
QY 1401 AlaLeuGlyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGlu 1420
Db 50042 CCGGCACCCAGCTCGAAGGCGCGCGCACCCGMAAGGCGACGACCGCGCGCGAAG 49983
QY 1421 AlaGlyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro 1440
Db 49982 CGCGGACGAAGA - - - - - AGTCGCGCGCGGTACGACC-GGCCGCTCCGGGATC 49936
QY 1441 LeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyr - - - 1459
Db 49935 CTGCGCAAGAAGCCC - - - - - GGCAACCCACCTTCGCGCACCGCGCGCGCAAG 49885
QY 1460 - - - - - AspThrGlyAlaSerThr - - - - - Gly 1467
Db 49884 CCGAAGGGNAGGCGACGACGCGCGCGCACCGCGCGCGCGCGCGCGCGCGCG 49825
QY 1468 SerLysLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProProVal 1487
Db 49824 GCGAAGAAGGCAAGACCGCTCCGCAAGGCGCGCACCGCTCCGCTCGCGTCCCAAG 49765
QY 1488 HisProLeuAspValMetAlaAspAlaArgAlaLeuGluAlaCysTyrGluGluSer 1507
Db 49764 GCGCGCGCGGGAAGTCCCGCAAGGCAAAAGATCAGCTCAGCGCGGGGATCGCGCACG 49705
QY 1508 LeuLysSerArgProGlyThrAlaSer - - - - - 1516
Db 49704 GCAGGTTCGCGGCGACCGACAGTACGGTTCGCTCGCGTCCCGCGCTTCCAGAA 49645
QY 1517 - - - - - SerSerGlyGlySerIleAlaArg - - - - - 1524
Db 49644 GGAGTACAGCCGCAAGTCTCGGGCGATGAAGTCCGACGCGCGCGCGCGCACCT 49585
QY 1525 - - - - - GlyAlaProValIleValProGluLeu 1533
Db 49584 TCGCGAGCGCGGTCTCAGACGCGGAGAGCGGTCAAGCGCGCGCGCGCGCGATGC 49525
QY 1534 GlyLysPro - - - - - ArgGlnSerProLeuThrTyr 1543
Db 49524 GCGCGCGGTCTCACCAGGTCAAGGTCAAGAGCGGAGAGAGACCGCGCGCGAAGA 49465
QY 1544 -GluAspHis - - - - - GlyAlaProPheAlaGlyHisLeuProArgGlyLys 1558
Db 49464 CCAAGACCAAGCCCGCGGTTCGCGCGCGAAGCGCGGTTCGCGCGCGCGCGGG 49405
QY 1558 rProValThrMetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLy 1578
Db 49404 CACCGA - CCGCGCGACCGCTCTGTCACGCGCAAGCGCGCGCGCGCGCGCGCG 49346
QY 1578 sAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSe 1598
Db 49345 ATCGGCACCAACGTCGCGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCG 49286
QY 1598 rThrValProGluHisProHisProHisProHisProHisProHisProHisLeu - - - - - 1614

Db 49285 CCGGTTCCCG-----CCGCTCCGCCACCCCGGACCACTGCGCCCGCCCTCG 49241
Qy 1615 -----LeuArgGlyValSerGlyValAspLeuTyArgse 1626
Db 49240 GCCGCGCGGTTCACGGGAGGTGCCACTGCCCGGCCCGCGG-----CGCG 49193
Qy 1626 rHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAsp-----1644
Db 49192 GCAGTCCG-----GCCACCGGCACCCCGCGCTGTCTCGCGGTTCGAGCGGAT 49145
Qy 1645 -----AlaAlaAlaAlaTy-TyrLeuProArgHis-----LeuAlaProAsp 1659
Db 49144 GGGCATGCGACGCGCGCTCCGCTCCGCGCGCCGATCAGGTGGAGCGCGGCGACTC 49085
Qy 1659 roThrTyProHisLeuTyProProTyTyLeuIleArgGlyTyPro--AspThrAlaAl 1678
Db 49084 CCGGCATCCCGCGCGCGAGTCCG-----GCACGGGGCTTCCCGCGGCACCTCGAG 49031
Qy 1678 aLeuGluAsnArgGlnThrIlelleAsnAsp-----1688
Db 49030 TAGAGCACGAAGAGAACACGCGTGTGACGCTGCGCTGACGAAGTGGGCGCGCAG 48971
Qy 1688 -----1688
Db 48970 CGTCGGGACATGACGCCCGACGACGAGACGAGTCACTCGACGACGCTCGAGT 48911
Qy 1689 -----TyIleThrSerGlnGlnMetHis---HisAsnThrAlaThrAlaMetAlaG 1705
Db 48910 CCTGGAGGACTCACCGGAGTCTTCAAGGCACACGACCTGCAACAGTGGCGCG 48851
Qy 1705 nArgAla---AspMet--LeuArgGlyLeuSerProArgGlu-----1717
Db 48850 GAAGCGCGGACCTCGCTACGAGTGGACGACCTCGCGCAGACCTGCGCTCCACCCA 48791
Qy 1718 -----SerSerLeuAlaLeuAsnTyAla 1725
Db 48790 CAACGTCACTCGCGCGCTGACCGCGCGCGATGGCGCAAGTGGCGAATCCATGGAGCT 48731
Qy 1726 AlaGlyProArgGlyIlelleAspLeuSerGlnValProHisLeuProValLeuValPro 1745
Db 48730 GTTGGCCCGAAGCCGACAGATGGCCACCAAGTCTCTACCG-CCGACAGACTCTCCG 48672
Qy 1746 ProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyLeuProThrAlaPro 1765
Db 48671 AGAGCGCGGAACCGCATGTTCAGC-----CCTACCGGCCCC 48633
Qy 1766 GlnProPheSerSerArgHisSerSerProLeuSerProGlyGlyProThrHisLeu 1785
Db 48632 -----TCCAGAACGCCACCGAGGACCGCGCTGCGCGTCCCGCGC-----48591
Qy 1786 ThrLysProThrThrThrSerSerSerGluArgGluArgAspArgGluArgAsp 1805
Db 48590 -----CCCGCTCCACACGAGNACTGAGAGGAGACACGACGTGAGCACAACCTCCC 48537
Qy 1806 ArgAspArgGluArgGlyLysSerIleLeuThrSerThrThrValGluHisAlaPro 1825
Db 48536 CGC-----48534
Qy 1826 IleTrpArgProGlyThrGluGlnSerSerGly-Ser-----SerGlySerSerG 1842
Db 48533 -----CATCCCGGCATGAACAACCGCGCTGGCGCAGCAGCAGCGGAGCGGCGG 48480
Qy 1842 yGlyGly-----GlyGlySerSerSerArgProAlaSer-----1853
Db 48479 CAACGGCTTCTCGCTCGCGCGCGGACCGCAAGCTCGCGCTCCCGCTCTTCCT 48420
Qy 1854 -----HisSerHisAlaHisGlnHisSerProIleSerProAr 1866
Db 48419 CAAGGAGGCGATGCACCTGCTCCGATCCACATCGAGCGCCACGCAACCGCGCCATCG 48360
Qy 1866 gThrGlnAspAlaLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyI 1886
Db 48359 CTTGTCGAGATGTGCGGCCAG-----48338

Qy 1886 eIleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSer-Serp 1906
Db 48337 -----GCCGAAGTCCAGCCCGCTTCACTCGATCGATCGAGACACCTCCACCGCCCTCAC 48282
Qy 1906 roVal-----ArgProAla-----AlaThrPhePro-----ProAlaThrHisCysp 1920
Db 48281 CGCGTGGCGAGCGGTCCCGCGAGATGGTCAACGCCCGCGCGCTCGACACCGCAGC 48222
Qy 1920 roLeuGlyGlyThrLeuAspGlyValTyProThrLeuMetGluProValLeuProL 1940
Db 48221 CCGCGCTTCAACGTGCGCACCGCGGAGTACC-----GCCCGGCTACGAGCGCTC 48168
Qy 1940 ysGluAlaProArgValAlaAlaArgProGluArgProAlaAspThrGlyHisAlaPheL 1960
Db 48167 CAACGGCTCCGGAGTCCAGCAGCCCAAGC-----CCGGCTTCTTACCGCG- 48124
Qy 1960 euAlaLysProProAlaArgSerGlyLeuGluProAlaSerSerProSerLysGlySerG 1980
Db 48123 -----GCACCGCGCGCGAGCTGAACAACCCCTCACCCACAGCAGCGGCGCG 48072
Qy 1980 luProArgProLeuValProValSerGlyHisAlaThrIleAlaArgThrProAlaL 2000
Db 48071 CACCGGACCG-----CCGGGTGCGGCCCTGCTGCTGCTGCTGCTGCTGCTGAAC 48027
Qy 2000 ysAsnLeu-----AlaProHisHisAlaSerProAspProProAlaProAlaSer- 2017
Db 48026 GGAGCAGCTGTTGAGCACCATGACCCAGCGCGCGCGCGCGCGCACACCAACGCGCGT 47967
Qy 2018 -----AlaSerAspProHisArgGluLysThrGlnSerLysp 2030
Db 47966 CCGCGCGCACGTCACTCACCGCGCGTCTCTGCGGACGACCACTCTACGCGCTCG- 47911
Qy 2030 roPheSerIleGlnGluLeuLysLeuArgSerLeuGlyTyHisGlySerSerTySerP 2050
Db 47910 -----CGCGCGCGGAGTGGCGCTGCGCGCGCGCGCGAGCTCGA 47877
Qy 2050 roGluGlyValGluProValSerPro-----ValSerSerProSerL 2064
Db 47876 CGACAGCGCTGCTCTACTCTCGCTCCGCGGAGTGGAGTGGGACACCTCGCTGCTGCT 47817
Qy 2064 euThr-HisAspLysGlyLeuProLysHisLeuGluLeuAsp-----2078
Db 47816 GTACGCGAAGACCAAGACCCCGACCGAGGACTGAGCGTCTGACCCCGCGCGAGCGCG 47757
Qy 2079 -----LysSerHis 2081
Db 47756 GATCCCGCTTCTGACGCGCGCGCACCTACTGTGGAACTGATCACTCCCGGCTTCAA 47697
Qy 2082 LeuGluGlyGluLeuArgProLysGlnProGly-----ProValLysLeuGlyGluAla 2100
Db 47696 CTGGTGGAGTCTGCGCCCGCGCGCGTCCCTGATGGGATGGGCGC---GCC 47640
Qy 2101 AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerProLeuLeu 2120
Db 47639 GCTCAC---CCGACGCGCGCGCT-----GGTCTCTGAGCT 47607
Qy 2121 GlnThrAlaProGlyValLysGlyHisGlnArg-----ValValThrLeuAlaGlnHis 2138
Db 47606 CACCGAGCACCGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGACCTACACCGATT 47547
Qy 2139 IleSerGluValIleThrGlnAspTyThrArgHisHisProGlnGlnLeuSerAlaPro 2158
Db 47546 CTTGCGCGGAGTGGCAGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47487
Qy 2159 LeuProAlaProLeu-----TyrSerPheProGlyAlaSer 2170
Db 47486 GTTCCGCG-CCGACCGCGCGACTTCGAGCGCGTCTGCTGCTGCGCGCGCGCGCGCG 47428
Qy 2171 CysProValLeuAspLeuArgProProSerAspLeuTyTyLeuPro-----2186
Db 47427 TGCCAC--CTTACCAGCGTGCAGCTGCGCGCGCGGCTTCCGACTTCCCGCTCGGACGCT 47370

2187 QY -----ProProAspHisGlyAlaProAlaArgGlySer-----ProHis 2199
|||||
47369 Db CAAGCTGCGCCGATGAAGGCTCCGGGCCCGCGGATCGCTGGTCCGCGCCAC 47310
|||||
2200 QY -----GlyLysArgSerProGluProAsnLysThrSerValLeu 2215
|||
47309 Db CTCGAGGACGTCGACCGCTCCGAGGACATCCGGGCT----- 47268
|||
2216 QY GlyGlyGluAspGlyLeuGluProValSerProProGluGlyMetThrGluProGly 2235
|||
47267 Db ---GTGGAGACTCGCTCTGCGC-----CCGGGCAAGCGCGCACCC----- 47229
|||
2236 QY HisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSer 2255
|||
47228 Db CGGTGTCGAGCTGCTGACATCCG-----GGCGAACCGCACCGCAT 47187
|||
2256 QY ArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPheSerLysLeu 2275
|||
47186 Db CGTCTGCTGCTCGCTCCCTCCCGCAAGAGATCCACTCCGCGCACCGCGATCTG 47127
|||
2276 QY ThrGluSerAsn-----SerAlaMetValLysSerLysLysGlnGluLe 2290
|||
47126 Db CTCGCGCTTCGGCATCGACGACCCCTCCGGCTCGTATCGAGACCGCATCCGCCA 47067
|||
2291 QY AsnLysLysLeuAsnThrHisAsnArgAsnGluProGlu-TyrAsnLysSerGlnProG 2310
|||
47066 Db GGGGCTGCTCTC-----CATCTACAAGACGACCCGCTGATAGGTCCGCAAGCCAC 47013
|||
2310 QY yThrGluLeuPheAsnMetProAlaLeuThrGlyThrGlyLeuMetThrTyr-ArgSerG 2330
|||
47012 Db CGTCGAAGACTGACGATGACGCC-----AAGGGCGCGCTCACCATCGCGCTCTG 46962
|||
2330 QY lnaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaLeuArgLysAlaL 2350
|||
46961 Db CCACGACGCAAGCCCGCGGATGCGGCTGTGGACCCGAGCGCGGCCCTCGCGCG- 46903
|||
2350 QY euMetGlyLysTyrAspGlnTyrGluSerProProLeuSerAlaAsnAlaPheAsnP 2370
|||
46902 Db -----GCATCACCGCGCGGCTCATCGCGCGCGGCAAGTC 46869
|||
2370 QY ro-----LeuAsnAlaSerAlaSerLeuProAlaAlaMetProLeuThrA 2385
|||
46868 Db CGTCTGCAGAACCTGATCTCTCGCGCGGAGAGCGCTCGCGGCTCTCTCTGGTCTCG 46809
|||
2385 QY laAlaAsp-GlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLys 2404
|||
46808 Db CGAGCTCAGGCGCGCATGAG-----CTCCGAGGCGGAGG-GCCGGG 46765
|||
2405 QY ValSerGlyArgProSerSerArg-----LysAlaLys 2415
|||
46764 Db TCGACTGTTCCCAAGGCGCGGTGGAGACGCTCTGTATGCTCACGCGCGCGCACCGG 46705
|||
2416 QY SerProAlaProGlyLeuAlaSerGlyAspArgProProSerValSerSer----- 2432
|||
46704 Db TCATGAAGTACCGGAGCGCATCAACGACGAGATGGCGCGCGGCACTTCGCCCTGAACA 46645
|||
2433 QY -----ValHisSerGluGlyAspCysAsnArgArg 2442
|||
46644 Db AGCCGTGCGCGCTCATCAACATCATCGTGGACGAGATCAACCGGCTGTCTCCACCGG 46585
|||
2443 QY Thr-----ProLeuThrAsnArgValTyrGluAspArgProSer--- 2455
|||
46584 Db ACGAAGCGATCGCAAGCCACCGCTACCTCATCGCGACATCCAGAACCGCGCGCA 46525
|||
2456 QY ---SerAlaGlySerThrProPheProTyrAsnProLeuLeuMetArg----- 2470
|||
46524 Db AGGTGCGCATCGCATCCGCTCGCGCTGCGAGTCCCTCACCTCAAGGACCTCGGTGACG 46465
|||
2471 QY -----LeuGlnAlaGlyValMetAlaSerProPro 2480
|||
46464 Db ACGACCGCATCGCGCAGGCGAAGTCGCGCGCTCTCTCATCGCACCGCTCTCT 46405
|||
2481 QY ProProGlyLeuProAlaGlySerGlyProLeuAlaGlyProHisHis-----AlaTyrAsp 2499
|||

46404 Db CCTCCA-----CCAAAGGAGATGGCCTGGAGC 46378
|||||
2500 QY GluGluProLysProLeuLeuCysSerGlnTyrGluThr 2512
|||||
46377 Db GCATCGCCCGCCGCGTT-----TCCAGATGGAACA 46345
|||||
RESULT 92
AY310323 138203 bp DNA linear BCT 06-DEC-2003
LOCUS Streptomyces sp. FR-008 heptaene macrolide complex synthesis gene
DEFINITION cluster, complete sequence.
ACCESSION AY310323
VERSION AY310323.1 GI:34766435
SOURCE Streptomyces sp. FR-008
ORGANISM Streptomyces sp. FR-008
KEYWORDS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 138203)
AUTHORS Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K.J., Lee, S.Y. and
Deng, Z.
TITLE Organizational and Mutational Analysis of a Complete
FR-008/Candidicin Gene Cluster Encoding a Structurally Related
Polyene Complex
JOURNAL Chem. Biol. 10 (11), 1065-1076 (2003)
PUBMED 14652074
REFERENCE 2 (bases 1 to 138203)
AUTHORS Chen, S., Huang, X., Zhou, X., He, J., Bai, L., Jeong, K.J., Lee, S.Y. and
Deng, Z.
TITLE Direct Submission
SUBMITTED (30-MAY-2003) Bio-X Life Science Research Center,
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,
China
FEATURES
source
1. .138203
/organism="Streptomyces sp. FR-008"
/mol_type="genomic DNA"
/db_xref="taxon:206662"
complement(574. .1950)
gene /genes="fscO"
CDS complement(574. .1950)
/genes="fscO"
/notes="putative FAD-dependent monooxygenase"
/codon_start=1
/transl_table=11
/product="FscO"
/protein_id="AAQ82549.1"
/db_xref="GI:34766436"
translation="MPSTCPAPTSIGFGCPHRRKRTSRGRIAVTPRCQKVRKGH
QAKWRVSMRGSSVAVVGGSIAGCAALAAASRGGAERTVFERADARLRDVGIGV
HDARYAELRDAGYADMEWPALNRRVWSVRDGAHGRAIGTQPPFRAYNWSLWS
ELRRVPETATYRAGAKBAVEQADGVTVRLADGEPFDLVIGADGYRSVVREMY
PTGAAAYAGYIGWGTSPDVSGLPSDGLDAHNITFPGGHMAVRI PDGSGGHLNWL
YAPPRIDGIDLPDRTPTSLPGRNLNAELTEHLRALVAEHPFPFAKILSTPAETVF
IQPILDYDPHYATDRNLIGDAASVARPHLGSVKALQDATALAAAVAGSWKEI
LEGYHAARGPVGVTAMVGLARRMGSAQVEETPDWSAMQAEFDAAWQEQNNSDRRSGF
GHSLSKSR"
complement(2264. .3037)
gene /gene="pabC"
CDS complement(2264. .3037)
/genes="pabC"
/notes="putative 4-amino-4-deoxychorismate lyase"
/codon_start=1
/transl_table=11
/product="PabC"
/protein_id="AAQ82550.1"
/db_xref="GI:34766437"
translation="MIELDGPAGPAPALASLALTNGHFTLLVNGRVRGLDIHLER
LIRDCRTLEDAALDPDRVKLARRAAPTGRATRVTVFDPALNIGNIAADARPGLV
TSRPAQGVKPLRVSVVHRRDLPEVKSGLCTLLRLRRQAQAGYDVLFTGPDGD
ILSGGTNNVGLVRDGEVWPGGLAGTTQLLRGHRHSRTGRPRGRLSRGLRD
QRVGVRPVTGIDIDREFPAHHSVTRLAEIYQALPGSPL"

```
gene
complement(3150. .3818)
/gene="fscRI"
CDS
complement(3150. .3818)
/gene="fscRI"
/notes="putative transcriptional activator"
/codon_start=1
/transl_table=11
/product="fscRI"
/protein_id="AAQ82551.1"
/db_xref="GI:34766438"
/translation="MDPAPADPAGSPERAAVLADGPDRAAGVACVACDPSLTITQVNOQ
EFERRGSSSELCSGRSDLVHPSVQOPLMHQFARMKDGKRRHSEIDAVIQAERTFA
STIPLNAAVGRGRTDVAAILVMNNAEEAGDADVMARKKLLSIDARILEGTA
GVSTIPLASRLYSRQGEYHVTGLLRKVPNRAALVSRAYSMGVLKVGTPPKVVE
DFIK"
complement(4377. .7205)
/gene="fscRII"
CDS
complement(4377. .7205)
/gene="fscRII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="fscRII"
/protein_id="AAQ82552.1"
/db_xref="GI:34766439"
/translation="MLLERTTEIALVDAAALGAATEGRSSVLTLTGLIGRSALLQRL
SCLADDDVRLRAHAAPMEQDFAGVVRQLPETLIGDSPEDARDRMDAHASFARH
VLAADAPACADQAIATAEVALHGLLSLLANVSADSRLLILVDDLOWSDVPSRLWTY
LAKRLHGLAVVNCALDGDPRSHHTLAREIREAGTQTLRPASLSLTATHELREHFG
EAGDGFVQACHASVNGDPLLSILVGTGLGRRRPLAETHAETARRRQLRRLAS
ILRTPAPVADLAAITAIGESDAPTLAAGLDSIGYAGALRALGALGALAAAPDEP
RFTHRSVRDAESTLTWQERWHEDEAALLYAAGSPAEQVAAQALMAVTVTRQPAVD
VLAADATLRGAPDPTAAGYLCRALDSPAAGVGRGLVELGTBERGDPPLACERH
IAOAMTLLPEPRVAALRI SPTALGPA PLTAVDLLRQAAEDLGPADGTLGTDRDLA
LRLEARLHGHCHDPHELASARHLKELPEPPVDTCGERELIALLGAATLGHPLA
SEVVRGLRVLDREPATSRLHSTLPLAITMTIAADSVEAVDSWLAVEHANEGRITP
ANALVIEQALVHLGRQAOREQSETALMGTEPAHQGLDTAATMTHTATVALESRD
SLAKVLRASRIRPQGLTIGTLRLLEAREIAAGDQGTALDSLSHSCGRQLAAAGH
VLFVPMRWPAWGVRRLGDIRTARSLAEBEYAEQWGAPOGVGRALRLLGLLEDGG
RGIGLLREAVTVLGRSANOELARALSTLAEIRGGGATELAREATLATSAGPL
AERAGAGHRSGPPATPEVLTPTERRVVVGLACRGLTQNETQIEIAGALGVSSRAVEKHLT
HAYRGLISGRRLVFLSRAEYAGTD"
complement(7210. .10254)
/gene="fscRIII"
complement(7210. .10254)
/gene="fscRIII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="fscRIII"
/protein_id="AAQ82553.1"
/db_xref="GI:34766440"
/translation="WYDLHMPPIELPILLDRDLAALSAVIGELSGRPVAVTVTCEP
GLGNDLLRWAAATADAGRVLSAHATPAEHEVRYGVVAQLLAENRALAPRLFTD
EQGGLGLNLAANDRPTLLVVRDQWDLPSLRWLEALARRLPAPFAVLLTST
GTALTPEWVGSLTGLATTIELALPPLTSGGTATVLRFAFGAPGDPFAFTEALAEAT
RGIPAVVDLDRFARAGYPRADRIDPLRTAEVVGDHATRALSDLRDPAADVARLE
ALQVGDLLDPLVCTLAGHPSVSESRALAAASGTLTLDGCHPRVQDAVVRARVL
EMPAADBAELIARAAGLAQRAVADDDQIADLLLLARVPDPAVDTLRGFTSALRG
RDLVAYLARALDEPLADEARIEFQLASVEMVTPAETAAERRLGLLIRATRPPTA
GLRARATDCLLGDTRAAHALAGIDSAPAEPRRGTGAPVPEEAPASTPT
HLVPHRYATGPGATGDRGAAAPDLPGTPRAHPDAPVHNPVAVAGVMAHEGVAGVRE
AVAEHELTVLCVASFLRHDDADLVLPVPAHPDAVHNPVAVAGVMAHEGVAGVRE
EVRLEARSAPVATGRPLVVRLLGCRALLTDDGDEAENHGLALLDDESHERAI
SVHAILTVRAELHGRGPDAARDLAAQAEPLDLRLHPLFLPYMLWALLSITDLQNG
HTHARARTEARPLPLAHESATTAQLLFARGVLTARDDEPHQREHFRACGRWLLRHG
CANPAVQDPLSLAAEAHTLGTDEEARLVHVEVRLARRWGAAPLGRAQLSLAVPE
ENVENLRAAVTTLISAASHTATYTRAVIELAAAREENERRTAVLQAPQTLVPPVAP
PCGPHVPSAVARPGPPVATLGTGRRTAKAPGHDAAVMHGLTAVERTATAALAAQCLGNR
EIAATELVATTVRELRLSUGYRKLIRNGRELRALVQEAESG"
complement(10298. .13315)
/gene="fscRIV"
complement(10298. .13315)
/gene="fscRIV"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="fscRIV"
/protein_id="AAQ82554.1"
/db_xref="GI:34766441"
/translation="MMLWAGDRSGAAGPQSGSAGPSPSLAPSDARESVPVPHRSRVP
TLVSGSRADVATVGYSCVRLAFGLIGLTKRGKRTSPLLVGARRALPALAADPCELOA
PAPPLVGRHOLDVLRHADARAGPRVLVLDGPGAGIKGTALLRAALAEADGPIAGMT
DCASTFVGLGYLAANIAMDGLVLVLDVHWCERSLRMLDFFLRADGPELLVJ
AARTGTGLTAPDALADLVNHLPLASLGLGLPLDTAEVAELTASFFPQDPLPSFVQ
LSVTGSPLEIVRLRLDRSAGLGGDDTGTARLAQAGKVKVAASFFVQBEHPQWVRQ
VARAVAVLDEDETRTLLAALGRVSVHVEAEVILRDAGLHPGHELSHDLVRAAVID
AVASGVAVLRRRAALLSDVGRSPEEIAQTLLVPGTDDMAVSVLRDAAAQAEQIG
ACEAAAYLERVREAPKPDVLSRLGLAETDPAVSVTLHHEHSLTDDVRAAAT
AVOLGTLCLAVQOSPQARALTEALDALDAELGPEPEPADRELTLTESALLIVGSD
KATLPDLIRTEGLTPQPGTTPAQROQLAMLSVLSAEGGAEETVGRARRALPAGV
PLGWSLLPTSLALSADENEAEEVLETVLRSGSDTAAVTVYVLAUSTSLKPLFENG
AVPDMAQAOTALEIILQOERWGDATMPHTYASVLTERTGDPGRALFALDAIKRPHLD
RFVMEYHVMARGRALAADCDLGDGALQVFGSCSMAEAGLTNPVLAPWMLTACILL
GEAGREEAGRAAAGHTRLAERWGTTRRAGYAAALARGAAAGTARTCALREAVALLAD
SPARSAGCLLLGLRALVEDGAVRSRHLREAVGLARPCGCVALARARDELIAAG
GNRREVTPLMLTGTERTVAGLVASGAGNRVRESLFTVTVTVLHLSVYRKLG
ARRGDTLEALREAGATAPAAERRPGHAKRNP"
13522. .14898
/gene="fscMI"
13522. .14898
/gene="fscMI"
/notes="glycosyltransferase responsible for mycosamine
attachment"
/codon_start=1
/transl_table=11
/product="fscMI"
/protein_id="AAQ82555.1"
/db_xref="GI:34766442"
/translation="MDSAARPIFLVSLPESGLNPLLVLAGELARRGVPLDWFATDEH
RREVEALSVDKSVSFASGEVVPDLSAVTWDDVREVTQSSRFKAHRAVVRQSYRP
ALQARKYRELEQVVDVVRPALMVVDVAVGVDLALARGIPYLVNVPFVASVNTLSH
PFGASVTPKSPVPNSGLPARMSVROKLANLTKWRTGLMFLHPDMAALREDAAIRK
ELGIAIPNATMYRDEAAVAVCSVAELDPFDPIDRVSLVGAVLPLPELPPDDDEVR
WLDQASVVMYMGFTITRLTREVAALVEVARRMSGTHQFLMKLPKQQLLPEAGSL
PDNLRVESVWPSQLDVLAHPNVSVFFSHGGNAYHEGVYFGKPVQVRLWVDCPDQAV
RGDRFGISLTLDKPHTVDPDDVVDVVDKLTTRVTSDFATREARLQALQRAAGRAAALD
VTGLLPAA"
14953. .16011
/gene="fscMII"
14953. .16011
/gene="fscMII"
/notes="putative aminotransferase responsible for
mycosamine biosynthesis"
/codon_start=1
/transl_table=11
/product="fscMII"
/db_xref="GI:34766443"
/translation="MAFTHPSRPAALDGRLELYVSDVSGWISSQGPYVRRFEAFA
EWNGVAGVACSSSTAALTLARALNIGDGEIVPEFTWVASAWVYTGATVPFVD
CGDLDNIDVTRIEKITPRTRAMPVHVYGRCCMDAVMDLALQYLVVVEDSFAHAG
VRVQDQSVLAFANKITTAGCGVCLTDPPRAEQALHARMAFTKDHSLFKKLYL
NYRMTAQGVALAQTREDEILATREIAREYADAGLKDLPGITLMPARDVLMYDLR
AERRELRHLARGIETRLFFKPMRQPGYLDVPMVPTLNAHRFSEDLGLPTHTGLT
AAQDEYITGAVDRDFRAG"
Alignment Scores:
Pred. No.: 0.000152 Length: 138203
Score: 662.50 Matches: 651
Percent Similarity: 30.22% Conservative: 247
Best Local Similarity: 21.90% Mismatches: 1082
Query Match: 5.01% Indels: 1004
Gaps: 1
US-09-522-753-5 (1-2517) x AY310323 (1-138203)
```

QY 43 GlnHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleLeuGlnProGln 62
Db 66773 CAGACCCCTGGCGA-----CCTGGTCGGGC-----CCACGC 66741
QY 63 ArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeu 82
Db 66740 GCGCGCGTCTCT----- 66729
QY 83 HisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGluPhe 102
Db 66728 -----CGGCCACGACGACCGCGTTCGA----- 66702
QY 103 IleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSerPro 122
Db 66701 -----CGCGCGCGCGC-----CTTCCGCGACTCGCTTCGACTCGCTGAC 66660
QY 123 LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr 142
Db 66659 CGCGTCAACTGGGACCG-----CCTCAGCGCGCGCACCGGTCTGGCGCT 66612
QY 143 GlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeuLeuLeu 162
Db 66611 GCCCGCACCTCTCTCTGACACCC-----GACCGTCCGCGCCCTCGC 66567
QY 163 ValProArgLeuSerLysGluGluLeuLeuIleGlnAsnMetAspArgValAspArgGlu 182
Db 66566 CGCCACCTCGA-----CGCGAACTCTCCGACCGCACCGCGCGCGCTCCCGCGCGC 66513
QY 183 IleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGlu 202
Db 66512 -----GCCCGACCGCTCCCGCGCGCGCGCGCGCGCGCGCGCTTCGCCAT 66465
QY 203 GluAlaAlaLysProProGluProGluLys-----ProValSerPro 216
Db 66464 CGTCGGCATGGCTCGCGCTCCCGCGCGACATCCACTCGCCGCGAGGACCTGTGGCGCT 66405
QY 216 ----- 216
Db 66404 GCTGGCGCAGCGCGCGCGCGCTCGCGCTTCCTCCCGCAGCAGCGTGGCGCTCGC 66345
QY 217 ProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArg-- 235
Db 66344 CCGCTCTACACTCGTCCCGAGGCGCGGAGCTCCGCGACCGCGCGAGGCGGCTT 66285
QY 236 -----LysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeu----- 248
Db 66284 CTTCCAGCAGCGCGCGACTTCGACGCGCGACTTCTTCGCACTCTCCCGCGAGGCGCT 66225
QY 249 -----GlyProGlnValGluLeuPro----- 255
Db 66224 CGCATGACCGCGCAGCGCGCTCTCTCGAGACGCGCTGGAGAGCTTCGACCGCGC 66165
QY 256 LeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAla 275
Db 66164 CGCATCGACCGCGCTCGTACGCGCGCGCGC-----CGCGCT 66123
QY 276 MetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLys 295
Db 66122 CTTCCGCGCGCTCTCTCCAGCGACTACTCAAGCGCGTCAACCCACGT----- 66075
QY 296 GlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysValGluArg 315
Db 66075 ----- 66075
QY 316 IleGluAsnAsnProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLys 335
Db 66074 -----CCCGCAGGAGTCCGCG----- 66057
QY 336 GlnPheProGluIleArgLysGlnArgGluLeuGlnGluArg-MetGlnSerArgVal-- 354
Db 66056 -----ATACGTCAACGCGCAGCGCGCAACGACGATCTCTCCGCGCGCTC 66009

355 -----GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisG1 371
66008 CTACACCTCGGCTGGAGGCGCGCGTACCGTCCGACACCGCGCTGCTCTCTCTCCCT 65949
371 uValser-----GluIleLeuAspGlySerGluGlnGluAsnLeuGluLysGlnMe 389
65948 CGTCGCGTGCACATGGCGCTGCGACGCGCGCGCAATCACCATGGCCCTCGC 65889
389 tArgGlnLeuAlaValIle-----ProMetLeuTyrAspAlaAspGlnGlnArg11 407
65888 GGGCGCGCTCACGTCATCTCTCCGCGCGAGATCATCTGGGACTTCTCCGCGACGCG-- 65831
407 elysPheIleAsnMetAsnGlyLeuMetAlaAsp-----ProMetLysValTyrLysAspAr 426
65830 -----GGCTTCGCGCGCGACGCGCGCTGCAAGCCCTTCGCGCAGCG 65790
426 gGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetG1 446
65789 CGCCGACGCGCACCGGCTTCTCGAG----- 65765
446 nHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaGluCy 466
65764 -----GGCTCGCGCTGATC-----CTCGTGGAGCGGCTCTCGACGCGCGC-- 65723
466 sValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgS8e 486
65722 -----CGCAACGCGCACCGAGTCTCGCTGCTGCTCGCGGCTC 65685
486 rTyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG1 506
65684 CGC-CGTCAACAGACGCGTGCACACGCGCTCAGCGCGCGCGCGCGCGCGCGCAG 65626
506 nGlnGlnGlnGlnProMetProArgSerSerGlnGluGluLysAspGluLysGluG1 526
65625 AGCGGTGATCCGCGACGCGCTCGC-----CGACGCGGTGTCGCGCGCGGAGTGCAGC 65569
526 uLysGluAlaGluLysGluGluLysProGluValGluAsn-AspLysGluAspLeuL 546
65568 TCGTCAGGCGCGACGCGCGCGCGCTCGCGCACCGATCGAGGCGCGCGCGCTCC 65509
546 eulysGluLysThrAspThrSerGlyLysAsnAspAsnAspGluLysGlu----- 562
65508 TCGCC-----ACGTACGCGCGCGACCGGACGCGCGCGCGCGCGCGCGCTTTC 65464
563 -----AlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGly----- 576
65463 TCGGCTCCGTCAAGTCCAACCTCGGCCACACCGCGCGCGCGCGGTGTCGCGGGGTGA 65404
577 --ArgArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaI 596
65403 TGAAGATGGTGTGGCGCTGCGCGCACCGCTGCTCCCGAGAGCTGCACATTGACGCGC 65344
596 leThrProGlnGlnSer---AlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgT 615
65343 CGACCGCGCACGTCGACTGGCGGAGTCCGCGGTGCGCTGCTGACCGAGCGCACCGCT 65284
615 rPthrGluGluGluMetGluThrAlaLysLys-----GlyLeuLeuG 629
65283 GGGCGGAGGCGCGCTCCGCGCGCGCGCGCGCGCTCTCTCTCTCGGATCAGCG 65224
629 luHisGlyArg----- 632
65223 GGCACCAACCGCCCATCTCTGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGG 65164
633 -----AsnTrpS 635
65163 CGAGCGGAGCGCCACCGCGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGG 65104
635 erAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPheA 655
65103 TGCTGTCCGCGCGCGCGCGCGCGCTGCGC---GGCAGGCGCAAGCGCGCTCTCACCC 65047
655 snTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnHisLysLeu----- 671

[illegible]

QY	2470	gLeuGlnAlaGlyValMetAlaSerProProGlyLeu-----	2484
Db	59446	-----AGGAGCGCCACCCCGCGGACCGGTGACGCCACCGACCGG	59406
QY	2485	-----ProAlaGlySerGlyProLeuAlaGlyPro-----	2495
Db	59405	CGCGTCCACCCCTGCGTCTCTCCGCGCTTCGACGCCGCTAGCGCCCGACCGCGGCA	59346
QY	2495	sHisAlaTrpAspGluGluPro 2502	
Db	59345	CGACTCTGCGCCACCTCACC 59324	
RESULT 93			
LOCUS	BHV1CGEN/c		
DEFINITION	Bovine herpesvirus type 1.1 complete genome.	linear	VRL 07-APR-2000
ACCESSION	AJ004801		
VERSION	AJ004801.1	GI:2653291	
KEYWORDS	complete genome.		
SOURCE	Bovine herpesvirus type 1.1		
ORGANISM	Bovine herpesvirus type 1.1		
REFERENCE	1 (bases 9702 to 103033)		
AUTHORS	Wirth,U.V., Fraefel,C., Vogt,B., Vlcek,C., Paces,V. and Schwytzer,M.		
TITLE	Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1 are 3' coterminal and encode a putative zinc finger transactivator protein		
JOURNAL	J. Virol. 66 (5), 2763-2772 (1992)		
MEDLINE	94219360		
PUBMED	1313901		
REFERENCE	2 (bases 103034 to 111027; 127191 to 135300)		
AUTHORS	Schwytzer,M., Vlcek,C., Menekse,O., Fraefel,C. and Paces,V.		
TITLE	Promoter, spliced leader, and coding sequence for BICP4, the largest of the immediate-early proteins of bovine herpesvirus 1		
JOURNAL	Virology 197 (1), 349-357 (1993)		
MEDLINE	94025583		
PUBMED	8212570		
REFERENCE	3 (bases 111028 to 114234; 123984 to 127190)		
AUTHORS	Schwytzer,M., Wirth,U.V., Vogt,B. and Fraefel,C.		
TITLE	BICP22 of bovine herpesvirus 1 is encoded by a spliced 1.7 kb RNA which exhibits immediate early and late transcription kinetics		
JOURNAL	J. Gen. Virol. 75 (Pt 7), 1703-1711 (1994)		
MEDLINE	94292919		
PUBMED	8021599		
REFERENCE	4 (bases 66900 to 96900)		
AUTHORS	Vlcek,C., Benes,V., Lu,Z., Kutish,G.F., Paces,V., Rock,D., Letchworth,G.J. and Schwytzer,M.		
TITLE	Nucleotide sequence analysis of a 30-kb region of the bovine herpesvirus 1 genome which exhibits a colinear gene arrangement with the UL21 to UL4 genes of herpes simplex virus		
JOURNAL	Virology 210 (1), 100-108 (1995)		
MEDLINE	95313343		
PUBMED	7793062		
REFERENCE	5 (bases 1 to 31444)		
AUTHORS	Schwytzer,M., Styger,D., Vogt,B., Lowery,D.E., Simard,C., Laboisserie,S., Misra,V., Vlcek,C. and Paces,V.		
TITLE	Gene contents in a 31-kb segment at the left genome end of bovine herpesvirus-1		
JOURNAL	Vet. Microbiol. 53 (1-2), 67-77 (1996)		
MEDLINE	97164286		
PUBMED	9010999		
REFERENCE	6 (bases 1 to 135301)		
AUTHORS	Schwytzer,M., Paces,V., Letchworth,G.J., Misra,V., Buhk,H.J., Lowery,D.E., Simard,C., Bello,L.J., Thiry,E. and Vlcek,C.		
TITLE	Complete DNA sequence of bovine herpesvirus 1		
JOURNAL	Unpublished		
REFERENCE	7 (bases 30801 to 67800)		
AUTHORS	Schwytzer,M., Vlcek,C., Lowery,D.E., Bello,L.J., Meyer,G. and Misra,V.		
TITLE	Gene contents in a 37-kb segment centered in the UL part of the		
	bovine herpesvirus 1 genome: the last gap		
	Unpublished		
	Accession# 278205		
	8 (bases 96901 to 99895)		
	Letchworth,G.J. and Kutish,G.F.		
	DNA sequence of the BHV-1 UL1 to UL3.5 genes		
	Unpublished		
	9 (bases 114235 to 122983)		
	Goltz,M., Buhk,H.J., Broll,H., Lewin,M., Mankertz,A., Boerner,B., Borchers,K. and Weigelt,W.		
	Nucleotide sequence of the HindIII O and K fragments located in the US region of the bovine herpesvirus 1 genome		
	Unpublished		
	Accession# Z98199		
	10 (bases 121402 to 123983)		
	Schwytzer,M.		
	Glycoprotein E and US9 genes of BHV1		
	Unpublished		
	11 (bases 1 to 135301)		
	Schwytzer,M.		
	Direct Submission		
	Submitted (08-SEP-1997) Institute of Virology, Faculty of Veterinary Medicine, University of Zurich, Winterthurerstrasse 266A, Zurich CH-8057, Switzerland		
	Location/Qualifiers		
	order(1. .4014,102918. .111027,127191. .135300)		
	/organism="Bovine herpesvirus type 1.2 (strain K22)"		
	/vifion		
	/mol_type="genomic DNA"		
	/strain="K22"		
	/db_xref="taxon:31519"		
	order(4015. .9664,15010. .22094,26636. .102917, 114235. .123983,135301)		
	/organism="Bovine herpesvirus type 1.1 (strain Cooper)"		
	/vifion		
	/mol_type="genomic DNA"		
	/strain="Cooper"		
	/db_xref="taxon:10323"		
	9665. .11962		
	/organism="Bovine herpesvirus type 1.1 (strain P8-2)"		
	/vifion		
	/mol_type="genomic DNA"		
	/strain="P8-2"		
	/db_xref="taxon:10324"		
	order(11963. .15009,22095. .26635)		
	/organism="Bovine herpesvirus type 1.1 (strain 34)"		
	/vifion		
	/mol_type="genomic DNA"		
	/strain="34"		
	/db_xref="taxon:31517"		
	order(111028. .114234,123984. .127190)		
	/organism="Bovine herpesvirus type 1.1 (strain Jura)"		
	/vifion		
	/mol_type="genomic DNA"		
	/strain="Jura"		
	/db_xref="taxon:31518"		
	<1. .472		
	/note="3'part of intron spliced out from transcript over covalently joined genome ends, leading to immediate-early RNA 1.5 and circ protein"		
	235. .430		
	/rpt_type=DIRECT		
	/rpt_unit="235. .248"		
	466. .1280		
	/product="LRL1.1 (circ)"		
	/note="transcribed from alternative late promoter"		
	/evidence=experimental		
	486. .1229		
	/note="immediate-early and late myristylated protein (vifion)"		
	/codon_start=1		
	/product="circ protein"		
	/protein_id="CAA06078.1"		
	intron		
	repeat_region		
	mRNA		
	CDS		

```

/db_xref="GI:2653292"
/db_xref="SPTREMBL:Q01342"
/translation="MGARASAPAAPPPAHAVILDLALSGGTIDLPGGDEAVFVSCPTT
RPVYHMRGRHTHTPVHFGVAYAILPCRKFWLYLMRGAVYGYEPTTGLHRLADS
LHDFLTATAGQORDLCLDVTVDLDAQMDPVTFTTPEILILEADPAPPPPPSARARS
TLRASRRPARFCFQHLLAEGSILDCSPFQNAAGCSLLLPACDSGDAAACPCDAGE
TARDCTADAAAPSPGALSRYSSVRSVFF"
gene
1259..1264
/genes="c1rc"
polyA_signal
1259..1264
/genes="c1rc"
gene
complement(1649..1654)
/genes="UL52"
/notes="synonyms: UL53, UL54"
complement(1649..1654)
/genes="UL52"
complement(1658..2860)
/genes="UL54"
complement(1658..2860)
/genes="UL54"
/notes="early BICP27 affecting mRNA processing (cell
nucleus)"
/codon_start=1
/product="UL54 protein"
/protein_id="CAA06079.1"
/db_xref="GI:2653293"
/db_xref="GOA:Q01354"
/db_xref="SPTREMBL:Q01354"
/translation="MADPEIATLSTASESDSLFGSDREDDDEAPSLAPALRSVVGQ
VRKRLGDEDEMPAPPGEGASGGPAEAPPARRARVRPRRRPRRRPAPAGE
QRSGPAAKEALATSHGGGGAARSIGSLRLARSLAEAAQRAATRAVTAFFAG
RLDLMRVONGFRAAGVSWAAVLDFAEQFVPEGRRTVTETLMFHGRDLRYMFVR
PAAOAAARLDLRSANLVDALASADECLTWCKFTATKNLRLTKDPIVATAGAVL
ENRLKLAPRLCYLREGSLPSLEELCAARLSLATCPASVMFMLARLSRAVRSAGE
CVPLEVTVDGAPFEETPGTCVAGLIDALDTHKQACDSMTCKLVANFTLVPVIMEK
YFYCNEIF"
gene
complement(3040..4038)
/genes="UL53"
complement(3040..4038)
/genes="UL53"
/codon_start=1
/product="glycoprotein K"
/protein_id="CAA06080.1"
/db_xref="GI:2653294"
/db_xref="GOA:Q65816"
/db_xref="SPTREMBL:Q65816"
/translation="MLLGGRTVNLALALLTAHLALALWVALAARQRCACVATARN
GSLRUEPSGAVTVGGANNATLADAPCRHAVVQHPGLDGDGAEALHGRVRAVAG
ARDCRAYLWCAQAGGLLWLVAFVYLRORRMFGLCRNDADFLSPGTYTLNYAAA
ALAAVGHGPTTKLARLMCELSARRALAVDFRLDPLGCARWRPRAALPLIAEGFARLG
ARTAAAGVGITHPCAAAYPLYLKIWAHVHVALPAGLELVSLLYRPRRGTCAGDG
GDGEGSIRKVCNCCSTLGLLVKALYLAIVGGVIALHYEHLRLRLQAQT"
complement(4013..7237)
/genes="UL52"
complement(4013..7237)
/genes="UL52"
/codon_start=1
/product="component of DNA helicase/primase complex"
/protein_id="CAA06081.1"
/db_xref="GI:2653295"
/db_xref="GOA:Q65817"
/db_xref="SPTREMBL:Q65817"
/translation="MDGPALEDALQILVATDCAIGFSLMLTGTGEEPRGAVVVS
NWPGRLAACLGSSPNDAADARPLVAFYLLGGPPGSSAATRLARLRLHLQOR
AGRARARRTPRQGRARAARPRGRHLCAARGHDHALLVATSQLAPRTGRADD
ARVVEGTATPSRLARPAGRGSLASLIHHEKTVAAVYRRLYSNSGATPFWFLSKFG
PGSKTLVATFTVFAERAGDAATDLOARDCLATYVAAPNPSPGLAFDPLVSFA
ALAAFCRSYGAAGAAGAPVVAARIADLAECVREYIDHRRSLKVADEPFA
VYVLAYFEGPNRRIQTEHLRAVTAPEPPDGGELAPRLAGVSRRLRERAVDAFPHVRA

```

Alignment Scores:
Pred. No.:
Score:

0.000153 Length: 135301
662.00 Matches: 656

Percent Similarity: 32.39% Conservative: 305
Best Local Similarity: 22.11% Mismatches: 1148
Query Match: 5.01% Indels: 870
DB: 14 Gaps: 138

US-09-522-753-5 (1-2517) x BHVICGEN (1-135301)

Qy	31	AlaArgThrHisThrAspValGlyLeuLeuGlu	---	TyrGlnHisHisSerArgAspTyr	49
Db	36388	GCAGCGCACGAGCTTAAAGTAGGCTTCGTAGAGCGCATCCGACGACACGACA	-GATCCC	36330	
Qy	50	AlaSerHisLeuSerProGly	-----	SerIleIleGlnProGlnArgArg	65
Db	36329	GGAAGGCGCGCTGCGCTCGCGCGCCAGCGCGGAGCTCGCAACCTCGCGTCGAAGG	36270		
Qy	66	ProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGluLeuHisLeuArg	85		
Db	36269	CCG-----	CCAGGAGGGCGCGACCCCTCTGTACGCGCGCTCGC	36231	
Qy	86	ProGluSerHisSerTyrLeuProGluLeuGly	-----	LySerGluMetGluPhe	102
Db	36230	CGCGCACCGCGCAAGCGTTGCTGCAGCGCTGCAGCGCGCGCTCCGCGCGCGCGCGG	36171		
Qy	103	IleGluSerLysArgProArgLeuLeuLeuLeuProAspProLeuLeuArgProSerPro	122		
Db	36170	CTTCGTGGAAGGCGCCAGCGCGTCCGCGCGCTCCG	36114		
Qy	123	LeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThr	142		
Db	36113	CTGTGAACG-----	CGCGCGCGCGCGCGCG	36078	
Qy	143	GlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeuGluLeu	162		
Db	36077	CCGCGCGCAGCCCGACAGCACCGCGCGCGCTCCAGCGCGCGCTCCGCGCGCGCTGG	36018		
Qy	163	ValPro-----	ProArgLeuSerLysGluLeuLeuIleGlnAsnMetAspArgValAsp	180	
Db	36017	CGCCCTCGTGTCTCTCGCGGAGCGG	35991		
Qy	181	ArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeu	200		
Db	35991	-----	35991		
Qy	201	GluGluGluAlaAlaLysProProGlu	-----	ProGluLysProValSerProPro	217
Db	35990	---CCGCGCGCGCGCGCGCGCGCGCACAAAGAGCTCCACAGCGCGCTCCACCGCGCGCGCGG	35934		
Qy	218	ProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysLys	237		
Db	35933	CCGAC-TCCAGCTCGCGCGCGCGCGCGG	35888		
Qy	238	AlaGluAlaAlaHisArgIleLeuGluGlyLeu	256		
Db	35887	GCAGCGCGCTCGATCGCTCCGCAAGGGCTCATGGCGCGCGCTC	35840		
Qy	257	TyrAsnGlnProSerAspThrArg	272		
Db	35839	-----	35839		
Qy	273	AsnGlnAlaMetArgLysLysLeuIleLeuTyrPhenylsArgArgAsnHisAlaArgLys	292		
Db	35791	CTGACGACCTCCACCGCGCGCGCGG	35735		
Qy	293	GlnTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLys	312		
Db	35734	CAG-----	35699		
Qy	313	ValGluArgIleGluAsnAsnProArgArgAlaLysGluSerLysValArgGluTyr	332		
Db	35698	GGCGGCGCGCTCCAGCGCTGGCGCGCGCGCGG	35666		
Qy	333	TyrGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnGluArgMetGlnSer	352		

Db 35665 -----CGCGCGCGCGTGTATCGCAGCAGGTTCTTGAG---- 35633
Qy 353 ArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluVal 372
Db 35632 ---GTTGGCAGCGCGCGCCCTCGTCGGTCCGCGCGCGCTCGCAGCACGACCTC 35576
Qy 373 SerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGlu----- 386
Db 35575 GCCGAG-----GATCGCTCGCGCGCGCGTGGCTCGCGCGCGCGCTTTGCCGC 35522
Qy 387 -----LysGlnMetArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGln 404
Db 35521 CTCTAGCGCTCTCGCGCGCGCGCTCGCTCGCGCGCGCT-----GCT 35477
Qy 405 GlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLys 424
Db 35476 CAGCATCTCGCTCTGAGCGCGCGCG-----AAGCGAGCGTGTCAAGCGCTCCAG 35420
Qy 425 AspArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPhe 444
Db 35419 CTGGCCAGCGCGCTCGCATCAGAGCTCGCTCGCGCAGCTCCAGCGCGCAGCGCTTC 35360
Qy 445 MetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAla 464
Db 35359 CAGCGCGCGCG-----CTCGCGCGCGCTGCGCAGTACGTACCTTCTTAGCGCGCT 35306
Qy 465 Glu----- 465
Db 35305 GATCTCGCGCGCGCGCGCGCGCTCGCAGCAGCTCGCGCGCGCGCGCGCGCGCG 35246
Qy 466 -----CysValLeuTyrTyrTyrLeuThrLys 474
Db 35245 CGCGCGCTCTCTCGAAGCGCGGTGCCCGCGGTAGAGCGCACAGTACTTGGCGAG 35186
Qy 475 LysAsnGlu-----AsnTyrLysSerLeuValArgArgSer 486
Db 35185 CCGAGGCGCGCGCGCGCGCTCCAGGTACGTTCAGCGCTCGTAGTAGCGAGCTC 35126
Qy 487 TyrArgArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 506
Db 35125 GCGCGCTCCGAGCGCGGACGTCAGCAGCGCGCGTGCATGTGATCAACCGCTCAG 35066
Qy 507 -----GlnGlnGlnProMetProArgSerSerGlnGluLysAsp 521
Db 35065 CTGSCCCACGCTCGCGCGCGGAGAGCGCGCAGCGCGCGCGCGCGCGCGCGCGCG 35006
Qy 522 GluLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAsp 541
Db 35005 GTCCACACCGCGAT---GCCGCGCAGCTCGCGCGCGCGCGCGGACATC-----GAT 34955
Qy 542 LysGluAspLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLys 561
Db 34954 CTTCGCGTTCTCGGCGCGTGGGTGAACCGCAGCAGCGCTCGATGCTCGCTCGCTCG 34895
Qy 562 -Glu---AlaValAlaSerLysGlyArgLysThrAlaAsnSerGlnGlyArg----- 577
Db 34894 CCGCTCGCGCTCGCTCCAGCGCGCGTGGCTCGCGCGCGCGCGCGCGCGCGCGTGA 34835
Qy 578 ----ArgLysGlyArgIleThrArgSerMetAlaAsnGluAlaAsnSerGlu---GluAl 595
Db 34834 GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGGAGTT 34775
Qy 595 aileThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerArgTr 615
Db 34774 CGCTCGAGCGGTCTCAGCGCGCGCGCGCAAGTCTCGCGCAGCGCTCTCTCGCGCTG 34715
Qy 615 pThrGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSe 635
Db 34714 CGCG-----GCCGCCCTCGCGCGCGCTCGCGCTCGCGCGCGCGCGCGCGACTTC 34661
Qy 635 rAlaIleAlaArgMetValGlySerLysThrVal-----SerGlnCysLysAsnPheTy 653
Db 34660 CTGCG 34601

Qy 653 rPheAsnTyrIlyLysLysArgGlnAsnLeuAspGluLeuGlnGlnHisLysLeuLysMe 673
Db 34600 CTGTGCGCCCAAGCGCGCGCGCGCGCATCGCGCGCGCGCTCGCTCGCTCGCTCGCT 34541
Qy 673 tGluLys-----GluArgAsnAl 679
Db 34540 CGCATCTCGCGCGCGCTCGCATCTCGCGCGCGCGCGCGCGCGCTCGCTCGCGCG 34481
Qy 679 aArgArgLysLysLysAlaProAlaAlaAlaSerGluGluAlaAla----- 695
Db 34480 CCGAAGTACTTCTCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCTGT 34421
Qy 696 -----PheProProValValGluAspGluMetGluAlaSerGlyValSe 711
Db 34420 TTTCGCGCGCATCAGCGCGCGCTCATGCGCTGCTAGTGTGCTCGCGCAGCGCAAG 34361
Qy 711 rGlyAsnGluGluMetValGluAlaGlu----- 722
Db 34360 CGCGCGCGAAGCTCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTG 34301
Qy 723 -----AlaLeuHisAlaSerGlyAsnGluValProArgL 734
Db 34300 CGCTCGCGCGCGCGCGCTACTGGCGCTGCGAGCGCTCAAAACGCGAACTCGCGCG 34241
Qy 734 yGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIleProSerPr 754
Db 34240 GAGCTGCGCGCGCGCGCGCGCTCGCGCGCTCGCGCGCGCGCTCGCGCGCG-- 34183
Qy 754 oHisThrGluAlaAlaLysAspThrGlyGlnAsnGly-Pro-----L 768
Db 34182 -----GCGGCCACCGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCG 34133
Qy 768 yAsProPro-----AlaThrLeuGlyAlaAspGlyProProProG 781
Db 34132 GCGCGCGCGCGCTCGCGTGTGCCCTCGAAACGCGCGCGCGCGCGCGCGCGCTCG 34073
Qy 781 lYPro-----ProThrProProArgArgThrSerArgAla---ProIleGluProT 797
Db 34072 GCCACGCGGTGTCACGACCGCGCGAAGGTGGCGCGCGCGCGCGCGCGCGCGCG 34013
Qy 797 hrProAlaSerGluAlaThrGlyAlaPro-----ThrProP 809
Db 34012 CGCAGCGCGCTCGGCTTCTCAGCGCTCGCGGTACCGCGCGCGCGCGCGCGCGCG 33953
Qy 809 roProAlaProProSerProSerAlaProProProValProLysGluLysGluG 829
Db 33952 GTCTCTGACCGCGCGCGCGCGCGCGCGCGCGCT-----CCTGC-AAGAAGCGCGCG 33900
Qy 829 luGluThrAlaAlaAlaProProValGluGluGlyGluGluGlnLysProProAlaAlaG 849
Db 33899 AGTCTCTGCGCTCTCGCGCGCGCGCTCTGCGCAGCGCTCGCGCGCGCGCGCG 33847
Qy 849 luGluLeuAlaValAspThrGlyLysAlaGluProValLysSerGluCysThr---- 867
Db 33846 -----GCG 33810
Qy 868 -----GluGluAlaGluGluGlyProAlaLysGlyLysA 879
Db 33809 GCACAGCGCGCTCGCGCTGTGTCAGCTCTCGCGCTCGCTCCCGCGCGCGCGCG 33750
Qy 879 spAlaGluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLysGlyG 899
Db 33749 GCGCGCGCGCGCGCGCGCTGCGCTCTAGCGCGCGCGCGCGCGCGCGCGCGCG 33690
Qy 899 lySerGlyArgAlaThrAla-----LysSerSerG 910
Db 33689 GCGCTCGCGCTCGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 33630
Qy 910 lyAlaProGlnAspSerSerAlaThrCysSerAlaAspGluValAspGluAlaG 930
Db 33629 CGGACCGCGCGCGCGCGCTCGCGCG-----TCTGCGCGCTCTCGCGCGCGCGCG 33576

Db 31676 AGTGCACTAGCGCGCCGCCACCGCGGCTCGGGCTGGCCACACGCGCGCAGAGCTAGGT 31617
Qy 1479 -----SerProGlyArgThrPhePro--- 1485
Db 31616 ACAGGTAGAGGCGCTCGGGGGCATCGCGCGAACAACGCTGGGCG-ACGTCCCGGCTC 31558
Qy 1486 -----ProValHisProLeu 1490
Db 31557 CCATCGGGTCGAAGAGTAGACCGGTCCGCGCGCGAAGACACCGAGCCCATGGCG 31498
Qy 1491 AspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSer 1510
Db 31497 CCACGATACACGCGCAGCGAGTCCGCGCGGTGTCAGACGCGCTCCACGAATTGC 31438
Qy 1511 ArgProGlyThr-----AlaSerSerSerGlyGly----- 1520
Db 31437 CGCGCTGGGACTGGGTGCTCAGGGCGCGCTCCGCGCGCGGTGTGAGAAGCGCGCACTCC 31378
Qy 1521 -----SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro 1536
Db 31377 CCATACACGCGCAAAACACGACGACAGCCGCTCCCGCGCCCGCTGGTGGTGATG 31318
Qy 1537 ArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArg 1556
Db 31317 CGGTTTGGCAGGTGCACGATCGAG-----CACATTCCGGG 31282
Qy 1557 GlySerProValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSer 1576
Db 31281 CGCGCCCGCTTTCGCGCGCCACGCGCGCCCTCAGAGAGACGCTCGATGGGTGCG 31222
Qy 1577 SerIleAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPro 1596
Db 31221 GCGCCACGCG-----GCGTCAGCGCGCGCGCGCGAAGCGCAGCGCAGA 31177
Qy 1597 HisSerThrValProGluHisHis-----ProHisProIleSerProTyrGlu 1612
Db 31176 AAGCAGACGAGGACCGAAGGATGCGCAGCGCTGCCCGCGCC-----AGCGCGCGGTGCG 31120
Qy 1613 HisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPhe 1632
Db 31119 TACTGTTGGAAACCCACAGCGGCGACCGCGACGTGCGCGCGGTGCGCGCGCGCGC 31060
Qy 1633 AspProThrSerIleProArgGlyIleProLeu----- 1643
Db 31059 GCGCGCGCAAG-----CGGACATGGTGCCTTCGCTCCACAGGGGTACTAGGGGCG 31003
Qy 1644 AspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThr-TyrProHis 1663
Db 31002 ACGCGCGCGCAGCGCGCTTATGCGAAGGCTCCCGCGCGCAGCGCGCGCGCTCA 30943
Qy 1663 sLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgG1 1683
Db 30942 CCGGGTGGCGCGCGGCGACTGGCGGCAACCGCATCGACTTCGGCG-----AGCAGGTC 30889
Qy 1683 nThrIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMe 1703
Db 30888 TTCCGTGTAAGC-----CACGGCCCGCGCGCGGT 30856
Qy 1703 tAlaGlnArgAlaAspMetLeu-ArgGlyLeuSerProArgGluSerSerLeuAlaLeuA 1723
Db 30855 TGCGCGCGCGCAGCAGCGCGGCGT-----CGCGCGCGCGCAGACGCGCGGCG----- 30809
Qy 1723 snTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGln----- 1736
Db 30808 -----GCTGACGGCGCGCGGTCTACCGCGCGCGCTCGCGCGCGCGCGCGGCGTC 30754
Qy 1737 -----ValProHisLeuProValLeuValProProThrProGlyThrProAlaThrAlaM 1755
Db 30753 CAGGAGCGCGCGCTCGCTCTC-----CCCGAAGCTCCGGGTAGCCCT----- 30707
Qy 1755 etAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerS 1775
Db 30706 -----CCACAGCGCGCAGCAGCGC-----TCTTCCGGCGCGCGCGCGT 30668

Qy 1775 exProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrSerSerSerS 1795
Db 30667 TTGCGCGCGTGGCGCGCGCGCGCGCTACCGCGGGCTCTCTCGGACCTCGTCTCC 30608
Qy 1795 luArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleu 1815
Db 30607 AGACGCGCGGAGCGCGCGCGCGCTCGCGACGCTCGCGC-----TCGCGCGCAC 30557
Qy 1815 euThrSerThrThrValGluHisAlaProIleThrArgProGlyThrGluGlnSerS 1835
Db 30556 TGTGCTGTAGCGCGCGCGCGCGCGCGCGCATCG--CGTCCAGCGCTGCAGAAAGCG 30499
Qy 1835 exGlySerSerGlySerSerGlyGlyGlySerSerSerSerArgProAlaSerHis 1855
Db 30498 GCCC-ACCTCGTGGAGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 30440
Qy 1855 exHisAla-HisGlnHisSerProIle-----SerProArgThrGlnAspAlaLeu 1871
Db 30439 CCAGCGCGCTCTCGCGCGCGCGCGCGCGCGAGTGGCGCGCGCGCGCGCGTGT----- 30383
Qy 1872 GlnGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGlu 1891
Db 30382 ---CGTGTGTGCG 30335
Qy 1892 ProSerLysProThrValLeuArgSerThrSerThrSerProValArgProAlaLeu 1911
Db 30334 -----CCTGCGCGCTT-----GTGCTGTCTCGCGCGCGCGCGCGCGTGA 30296
Qy 1912 ThrPheProProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThr 1931
Db 30295 AGG---CCGCCACGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30263
Qy 1932 LeuMetGluProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgPro 1951
Db 30262 ---CGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30212
Qy 1952 ArgAlaAsp-----ThrGlyHisAlaPheLeu 1960
Db 30211 AGCGCCACCG 30152
Qy 1961 AlaLysProProAlaArgSerGlyLeuGlu-----ProAlaSerSerProSerLys 1977
Db 30151 GCG 30092
Qy 1978 GlySerGluProArgProLeu-----ValProProValSerGlyHisAla 1992
Db 30091 CACGTGCG 30032
Qy 1993 ThrIleAla-----ArgThrProAlaLysAsnLeuAlaProHisHisAlaSer 2008
Db 30031 CCTGCG 29978
Qy 2009 ProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSer 2028
Db 29977 -----CG 29960
Qy 2029 LysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyr 2048
Db 29959 -----CGGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29909
Qy 2049 SerProGluGlyValGluProValSerProVal---SerSerProSerLeuThrHisAsp 2067
Db 29908 CTACCGTGTGCG 29861
Qy 2068 LysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyLeuArg 2087
Db 29860 -----TCGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29834
Qy 2088 Pro-----LysGlnProGlyProValLysLeuGlyGlyGlyGlyGlyGly 2102
Db 29833 CCACAAACG 29774

Qy	1093	rgProValLeuProArggProThrIleSerAenProProProLeuIleSerSerAlaL	1113
Db	9234	--CGGCCCATCCGCGAAGCCTTGCACCGCCATCGCGCCAAACCCCGCTCGCGCA	9177
Qy	1113	yHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnL	1133
Db	9176	AAATCTCCAGCA-----ACAAATCCGGGTGCGCCCATCATCCGTCAC	9138
Qy	1133	euHisValProTySerGluHisAlaLysAlaProValGlyProValIleThrMetGlyLeuP	1153
Db	9137	ACCACCGGCCAACGCCATGTGCACACTCGCCACCCCGCAACGCCTGCG-----	9091
Qy	1153	roLeuProMetAspProLysLeuAlaProHisSerGlyValLysGlnGluLLeuS	1173
Db	9090	-----CCGCGTGTGCGGGCCACCAGCGACGACGAAACACGCGGTGC	9048
Qy	1173	erProArgGlyGlnAlaGlyPro-ProGluSerLeuGlyValProThrAlaGlnGluAla	1192
Db	9047	GATGAGAGCAAGAGACCTCCAGCCCGCGGTGTACGAGATCGCGCCGAGACGACGT	8988
Qy	1193	SerValLeuArgGlyThrAlaLeuGlySerVal-----ProGly	1205
Db	8987	CGCGCATTTGCCGTGAGCAGATAGCTTCGTACTCTCCGGGGCTCGTGCAGCCGGA	8928
Qy	1206	GlySerIleThrLysGlyIleProSerThrArgValProSerAspSerAlaIleThrTyr	1225
Db	8927	GGCGTACTCTGTATGGTTCAGCCCGAC-----GTATAC	8895
Qy	1226	ArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArg	1245
Db	8894	CCCGGTGCG-----GCTGCGGTGCAGCGCA	8871
Qy	1246	IleIleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLys	1265
Db	8870	GTG-----CGGTCGATACCCGCCCGCTCGATCGCTCCATCCCTCCCGGCA	8832
Qy	1266	Gly-----HisVal-IleTyrGluGlyLysGlyHisValle	1278
Db	8831	GGTCTCCAGCAACAGCCGCTGCTCGCGATCCATCGCCACCGCTCACGCGCGCAATCCC	8772
Qy	1278	uSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgSer-SerS	1298
Db	8771	GAAGAACTCCGCGTCGAACCTCCGCGCGTGTGGAGAAAACCCCTCCCGCATATGT	8712
Qy	1298	erGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgV	1318
Db	8711	CTTCCC CGCACCCCGCTCGGTCGATAGACAGGAATCCAGGTCCAGCCCCCGTCTG	8652
Qy	1318	alGlyArgAlaIle-----SerSerAlaSerIle-----	1327
Db	8651	GGGCGCGCGCCATATGACGTACGCTCCTCGGTACCGACTTCACAGCGCTCAGGGGT	8592
Qy	1328	-----GluGlyLeuMetGlyArgAlaIleProProGluArgHisSerP	1342
Db	8591	GCCCGCCCCACCGGGAAACGGCAGCCATCGCGACGATGACCAACCGGGTCGTCTGCGC	8532
Qy	1342	roHisHisLeuLysGluGlnHisIleArgGlySerIleThrGlnGlyIleProArgS	1362
Db	8531	CACGG-----	8527
Qy	1362	erTyrValGluAlaGlnGluAspTyrLeuArgGluAlaLysLeuLeuLysArgGluG	1382
Db	8526	-----CGACGGGCGAGTGGCGGTCTATCGCGGGCG-G	8497
Qy	1382	lyThrProProProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaL	1402
Db	8496	GCGCGGGGCTCCGGCCCGGTACGCGGTGAGCAGCAGTCTGGCCAGCGCGGTCTGGGC	8437
Qy	1402	euglyProLeuLysLeuLysProAlaHisGluGlyLeuValAlaThrValLysGluAlaG	1422
Db	8436	TGGGGTGGTCCGAAGATG-----ACGGTGGCGGGTAGCT	8404

[illegible]

Db	7384	GC	CGCGGC	CGCGACAC					CGCGCGCGGTAGATCCC	-----	7350											
Qy	1669	Leu	Ile	Arg	Gly	Tyr	Pro	Asp	Thr	Ala	Ala	Leu	Glu	Asn	Arg	Gln	Thr	Ile	Ile	Asn	Asp	1688
Db	7349	-----	CGAGGC						CGAGGC	-----	7344											
Qy	1689	Tyr	Ile	Thr	Ser	Gln	Met	His	His	Asn	Thr	Ala	Met	Ala	Gln	Arg	Ala	Asp	1708			
Db	7343	-----	GGC	AC	CGCGAC	CGC	CTGGT	CCT	CGCGT	CGCGCGCGGAGGAC	7302											
Qy	1709	-----	Met	Leu	Arg	Gly	Leu	Ser	Pro	Arg	Glu	Ser	Ser	Leu	Ala	Leu	Asn	Tyr	Ala	1725		
Db	7301	GGC	GGCAGC	AACT	CAAGG	CGCGT	CGGT	TCGG	AGT	TCGGG	CAGCCT	TC	CGG	GGT	CT	7243						
Qy	1726	Ala	Gly	-----	Pro	Arg	Gly	Ile	Ile	Asp	Leu	Ser	1735									
Db	7242	GTG	GGG	GGCTGT	AGGT	TC	CGCGCT	CGCGT	CAGCT	CCGCT	GTCT	CGCGCAGATCGATCA	7183									
Qy	1736	Gln	Val	Pro	His	Leu	Pro	-----	Val	Leu	Val	Pro	Thr	Pro	Gly	Thr	Pro	Ile	Thr	Ala	1754	
Db	7182	AGC	GGCCCCAGAT	TC	CGGAGT	TC	AGGCGCCG	CGCT	CGT	CGA	AGCCCA	CTGGG	7123									
Qy	1755	Met	Asp	Arg	Leu	Ala	Tyr	Leu	Pro	Thr	Ala	Pro	Gln	Pro	Phe	Ser	Ser	Arg	His	Ser	1774	
Db	7122	CCC	CGCGGT	CCT	CGCGCT	CT	CCC	CGGGA	ACC	CAGCG	CAC	CGGCC	CT	CGCGT	CAGCGCC	7063						
Qy	1775	Ser	Pro	Leu	Ser	Pro	Gly	-----	1780													
Db	7062	AG	ACCT	TGGCGCGCGG	ACCG	CCGAT	TC	CGAGAC	CGCT	CGAGAG	CGCTAT	CGACGTCC	7003									
Qy	1781	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1786									
Db	7002	TG	ACCGCGT	CGTGG	AGGG	CAC	CAGAC	CCGAG	ACCGAG	CAGC	ACGCGT	GTCTCT	CTCCC	6943								
Qy	1784	His	Leu	Thr	-----	-----	-----	-----	-----	-----	-----	-----	1786									
Db	6942	CCG	CTGACT	CCAT	GAGGG	CT	TGAGC	AGT	TCGG	CAC	CTCGGC	ACGCGCGCCGATGG	6883									
Qy	1787	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1787									
Db	6882	GC	ACT	CAAGCGCGG	CGAC	CGCGCT	CGGAC	AAAC	CGT	TGT	CAGT	CCGTCGA	6823									
Qy	1798	Arg	Asp	Arg	-----	-----	-----	-----	-----	-----	-----	-----	1806									
Db	6822	GG	ACT	CGCGT	CGGT	CGGT	CGGGGT	TC	ACG	AGGAGG	CAGCCAT	TGAGT	GAGGGC	6763								
Qy	1807	Asp	Arg	Glu	Arg	Glu	Ser	Ile	Leu	Thr	Ser	-----	Thr	Thr	Thr	Val	Glu	His	Ala	Pro	1825	
Db	6762	GG	ACGGGG	GAG	ATT	CGGT	CAGG	CT	TTCC	AGGT	TTC	ACG	GT	TTC	ACG	GT	TTC	ACG	GT	TTC	6706	
Qy	1826	Ile	Trp	-----	-----	-----	-----	-----	-----	-----	-----	-----	1827									
Db	6705	---	TGGGG	CT	TCGGT	CGT	CGT	CGCGGGG	TG	TGGGGT	GTGT	CGAGC	CA	GTA	GT	AGT	GGT	6649				
Qy	1827	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1827									
Db	6648	G	T	G	T	T	G	A	G	G	G	T	A	G	G	T	G	T	G	G	T	6589
Qy	1828	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1839									
Db	6588	AG	G	T	A	G	G	T	G	G	C	C	T	G	G	G	T	G	G	G	T	6529
Qy	1840	Ser	Ser	Gly	Gly	Gly	Gly	Ser	-----	-----	-----	-----	Ser	Ser	Arg	Pro	1851					
Db	6528	T	G	T	C	T	C	T	G	C	G	C	G	C	A	G	G	T	C	C	G	6469
Qy	1852	Ala	Ser	His	Ser	His	Ala	-----	-----	-----	-----	-----	1857									
Db	6468	T	C	T	C	T	C	G	A	T	G	A	C	G	A	G	T	G	A	C	G	6409
Qy	1858	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	Se	1861								

[illegible]

```
QY 2037 uLeuArgSerLeuGlyTyrHisGlySerSer----- 2047
Db TCAGCGGTGTGGGGGCTTCAGAGGATCAGATGGCGGTGGTACCACTGACACCAAAAGCG 5225
QY 2048 -----TyrSerProGluGlyValGluProValSerProValSerProSerLe 2064
Db GAGACCGCGCCGCGCGCGGTCCACCCCGCGGCATCCCGCGCGCTCGGTGAGCAAC 5165
QY 2064 uThrHisAspLysGlyLeuProLysHisLeu-GluGluLeuAspLysSerHisLeuGluG 2084
Db CACACACACCGACCGACGAGTCAATCGCGCTCGGCTC-----ATCCACATGCAA- 5112
QY 2084 lyGluLeuArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaHisLeuP 2104
Db CCGTCCGCGGCAACCA-----CCCATCCGCGCATCCGCAACACCATCTTG 5069
QY 2104 roHisLeuArgProLeuProGluSerGlnProSerSerProLeuLeuGlnThrAlap 2124
Db ATCACAC-CAGCCACCCCGCAGCGCGCTGAGCATGACCAATATTCGACTTCAACGACCC 5010
QY 2124 roGlyValLysGlyHisGlnArgValThrLeuAlaGlnHisLeuSerGluValIle 2144
Db CA-----ACCA 5004
QY 2144 hrGlnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeu 2164
Db CACCGGAGCGCCACCATCAGATCCGCGCCATAGTCCCGCAGCAGCGCTGCGGCTCGAT 4944
QY 2164 yrSerPheProGlyAlaSer-----CysProValLeuAspLeuArgArgProPro 2181
Db CGGATCACCCAGCGCGGTCCCGTCCCATCGGCTCCACCATCCACATCCGCGCCACCGT 4884
QY 2181 erAspLeuTyrLeuProPro-----AspHisGlyAlap 2193
Db CAACCGCGCATCCCAACCGCTCCGAGCAATCACCGCTCTGGAGCGACCATCGCGC 4824
QY 2193 roAlaArgGlySerProHisSerGluGlyGly-----LysArgSerProGluProAsnL 2211
Db CGTCAACCATTCGACGACCGCTCTGATTCACCGCGCGCCACACACACCGCGCAAC 4764
QY 2211 ysThrSerValLeuGlyGlyGlyGluAspGlyLeuGluProValSerProProGlu---- 2229
Db CGGATGCCCAACCGCGCGCATCCGACACCGCTCCACGACGACACCGCGCGCCCTC 4704
QY 2230 -----GlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArg 2248
Db GCGCCCAAGCGTCCATCCCGCCCATCCCGGAAC-GCC-----TTGACCGCC 4657
QY 2248 spGlyGluGlnThrGluProSerArgMetGlySerLysSer-----ProGlyAsn 2265
Db CATCGCGCGCAACCGCGTCCGCGAATACTCCAGACATCCCGCGCGCGCATCA 4597
QY 2265 hrSerGlnProProAlaPheSerLysLeuThrGluSerAsnSerAlaMetValLys 2285
Db CCGTCAACACCGCGC-----AACGCCAT-GTCACACTCG 4562
QY 2285 erLysLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArg----- 2299
Db CCACCGCGCAACCGCTGACCGCGCAGATGCAACGCCACCAACGACGACGCGGTA 4502
QY 2300 -----AsnGluProGluTyrAsnI 2306
Db TCACCGGTACGCGCGGACCTCCAGCCCGCGGTAGCGCCACACCGCCGACACCA 4442
QY 2306 leSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMet 2326
Db CTCCCGACGTACCGGTGAGGAGGTAGCGCTCGCTCCGCTCCGCTCCGCTCCG 4388
QY 2326 hrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIle 2346
Db AGCCGAGGTCCGTATTCCTCGCGCATCACCGCGCGGTAGACCGCGGTGCGGCTGC 4328
```

```
QY 2346 leArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuSerAlaA 2366
Db AGCGAGT-GCG-----GGTCGATACCGGCC---GCTCGATCGCTCCCGAGG--- 4282
QY 2366 snAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaA 2386
Db -----TCTCCAGCAACAGCGCTGCTCGGATCATCGCCACCGCTCCACGCGCGA 4230
QY 2386 laAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValS 2406
Db -----ACTCCGCG-----CATCGCGAANAACC 4182
QY 2406 erGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspA 2426
Db CCCTCCCGCACATACGATTCCTCCGCCACACCGCGCTCGGTACGAACC--- 4129
QY 2426 rgProSerValSerValHisSerGluGlyAspCysAsnArgArgThrProLeuT 2446
Db -----CCTCCACATCCCGACCGATCACCGGGAACCGCGTCCACCCCT--- 4075
QY 2446 hrAsnArgValTyrGluAspArgProSerSerAlaGlySerThrPheProTyrAsn 2466
Db -----CCACCACCAACCGCCACACGCT---CCGCGGAC 4041
QY 2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProPro-----ProProGlyL 2484
Db -----CCCCACCGGGAACCGGCACCCCATACCCACCGCGATCGCTCTGGC 3987
QY 2484 euProAlaGlySerGlyProLeuAlaGly 2493
Db TTGAAAGCTCTTCTGCGCGGTGCGAGGA 3958

RESULT 95
AE007164/c 15311 bp DNA linear BCT 27-APR-2001
LOCUS Mycobacterium tuberculosis CDC1551, section 250 of the
DEFINITION complete genome.
ACCESSION AE007164 AE000516
VERSION AE007164.1 GI:13883468
KEYWORDS Mycobacterium tuberculosis CDC1551
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
REFERENCE 1 (bases 1 to 15311)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 15311)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source 1. .15311
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="Genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
complement(114. 1662)
/gene="MT3613"
/note="This region contains an authentic frame shift and
```


8744	AGCAGACCGAGGCGCGGCCAGCCGATGACGTGATCGGGGACACGTGACGAGCGGTTCTCT	8686
1522	lealaArgGlyAlaProValileValProGluLeuGlyLysProArgGlnSerProLeuT	1542
8684	GATTGCGCGGTGCGG-GCTATGGCGCGCGGCTGGGGATACCGGTGGGTTCGCGGTG	8626
1542	hrTyrGluAspHisGly-AlaProPheAlaGly-----	1552
8625	TCGTCAATCT--GGCTGGACCACCTCGCGGGGAGCGCCCGCGCACCATCGCGCGCGC	8569
1553	-----HisLeuProArgGlySerProValThr	1561
8568	TGGCCCATCGGACTAGACCGCCGCTGCGCCCTTGC CGCGCTCCGCGTCCGCGC	8509
1562	MetArgGluProThrProArgLeuGlnGlySerLeuSerSerSerLysAlaSerGln	1581
8508	CCGCGCGCGCGGTGCA-----CCGGCTGGTCTTACACCGTCTCCGGCGCCACCG	8455
1582	AspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPro	1601
8454	CCGCTGCGGTGTCTCCTTGC CG-----CCATTCCAGCCGCTGCGCG	8413
1602	GluHisHisProHisProLieserProTyrGluHisLeuLeuArgGlyValSerGlyVal	1621
8412	CCGTTCCACCAACACCG-----	8395
1622	AspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIle	1641
8394	-----CCGGTCCGCCATCGCGCGGTGCGCCCGAGCGGTG	8356
1642	ProLeuAspAlaAlaAlaIaTyrLeuProArgHisLeuAlaProAsnProThrTyr	1661
8355	CCACTGCG-----TTGCGCGCGGTGCCACCCCTTG	8326
1662	ProHisLeuTyrPro-----ProTyrLeuIleArgGlyTyrProAspThrAla	1677
8325	CCGCGGTGTGCGGCGCGCCAGCACCAACCGGCTCCCGGGGGCACTCCGGCTCGCG	8266
1678	AlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHis	1697
8266	-----	8266
1698	AsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGlu	1717
8265	-----TCGCG-----	8360
1718	SerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnVal	1737
8259	-----CCATTGCCCGCGGTG	8245
1738	ProHisLeuProValLeu-----ValProProThrProGlyThrProAlaThrAlaMet	1755
8244	CCGCGGATACCGCGGCCACCGGTGCCCGCTTGCATTTGCTCCCGCGCGGTG	8185
1756	AspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSer	1775
8184	CCGCGTTGCGCATTTGCCGCGCACCG-----	8155
1776	ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlu	1795
8154	CCGCTGCGCG-----GTGCGCGCGTGC GC--GCCGATCCGCGCAC	8114
1796	ArgGluArgAspArgAspArgGluArgAspArgAspArgGluArgLysSerIleLeu	1815
8114	-----	8114
1816	ThrSerThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSer	1835
8113	-----CCGGGGTTCGCGAGCGCGCC	8093
1836	GlySerSerGlySerSerGlyGlyGlySerSerSerArgProAlaSerHisSer	1855
8092	GGTGC CGCGGTGCGCGGTGCGCGCG-----GCCCGCTCCCGCGCGCGGTG	8039

Qy	1856	HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnArgPro	1875
Db	8038	GCGGGTCCGCACATCGCGCATTTGCCGGCGCAGCACCCCGGTGCGCGCGCGCGAC	7979
Qy	1876	SerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaVal-GluProSerLysPr	1895
Db	7978	CTCGCCGCCCTTCGCGCGCGGTACC-----GCGGTTCGCGCGGTTCGCCATCATCGCGCC	7925
Qy	1895	o---ThrValLeuArgSerThrSerSerProValArgProAlaAlaThrPhePr	1914
Db	7924	AGTGTCGCCGCCAAGGTGCGATGCCCCGCGCGCGTTCGCGCGGTGCGCGCGTTCGC	7865
Qy	1914	oProAlaThrHisCysProLeuGlyGlyThrLeuAspGlyValYrProThrLeuMetGl	1934
Db	7864	GCGGTGCCCGGTTCGCGCGGGTCTCAGTAGTGGCGGTGGCGCGCG---ATGGC	7808
Qy	1934	uProValLeuLeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAs	1954
Db	7807	CCCC-----TTGCGCGCATTCGCGCGCGCGCGCTTGC CGCGCGGTGCGTGGCGCC	7754
Qy	1954	pThrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluProAlaSerSe	1974
Db	7753	AGCG-----GTCAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAAGCC	7709
Qy	1974	rProSerLysGly-----se	1979
Db	7708	ACGGTCCGCGCGTTCACCGTCACCGCGCGCTGCGCAGTCATGTTGCGGTATTATC	7649
Qy	1979	rGluProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAl	1999
Db	7648	TGCTCCGTTCGCGCGGTACCGCGCTCACCGCGCATGCCACCGTGC CGCGGTGCGCGC	7589
Qy	1999	aLysAsnLeuAlaProHisHisAlaSerProAppProAlaProProAlaSerAlase	2019
Db	7588	AGTACCGGTGAGCGC-----CCGTGCGCGCGGTGCGCGCGCGCGCGCGCGC	7541
Qy	2019	rAspProHisArgGluLysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuAr	2039
Db	7540	GCCGCG-----7534	
Qy	2039	gSerLeuGlyTyriHisGlySerSerTyrsrProGluGlyValGluProValSerProVa	2059
Db	7533	-----TTCCGCGCG------GCACCGCGCTGCGCGCT	7508
Qy	2059	lSerSerProSerLeuThrHisaspLysGlyLeuPro-----LysHisLeuGluGluLe	2077
Db	7507	GCGCGTCCCGTGTTCGCGCCCATGCGCGCGTTCGCCCGGTGCGCGCGTTCGCGCGGT	7448
Qy	2077	uAspLysSerHisLeuGluGlyGluLeuArgProLysGlnProGlyProValLysLeuGl	2097
Db	7447	GCGCGCGTTCGCGTTCGCGCACCGCGGTTCGCGCCCATGCGCGCG-----7396	
Qy	2097	yGlyGluAlaAlaHisLeuProHisLeuArgProLeuPro-----2110	
Db	7395	-----ATCGCGCGTTCGCGCGTTCGCGCGGTTCGCGGTTCGCGGTTCGCGCGC	7343
Qy	2111	-----GluSerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyVa	2126
Db	7342	GTTGCGCGCGTTCGCGCGGTTCGCGCAGCGCGCGGTTCGCGCGGTTCGCGCGCGCT	7283
Qy	2126	lLysGlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAs	2146
Db	7282	GCCC-----7279	
Qy	2146	pTyrrThrArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeutyrrSerPh	2166
Db	7278	-----CCAGCGCGCGCGCGCACCGTTAGCACCAACCAACCGCGTTCGC	7235
Qy	2166	eProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeutyrrLeuPr	2186
Db	7234	GCCAGCCCGCGCGGTGCATTCGCGGAATTCGCGTTCGCGCG-----TTACC	7187


```

/product="AmphdIII"
/protein_id="AAK73500.1"
/db_xref="GI:14794892"
/translation="MPKRALITGITGDSGLAEHLISQGVYWGILGOANPKFRV
SRLASLFDVGDLMOGSLVSDVQDEYVNLCAISFVPMNSWQOABLVEYNGWG
VLRLSARVNSGLSMRSTAGTSGQIRFQASSENFGKVAETPQREITLHFHSPYG
AAKAGHFITRNRYRESFGVAVSMGLFNHSEPRGQEFVTRKTLISLAVARIKLGLQDKL
ALGNMADYVAGVYVAVHMLMLQODAPDDYVIGTGBMHTSLAVARFAFHEGLDWM
KDYVVDPLVRAEVEVLCASSKAQAQLGKWPQVDFQLMRMMVMDADLASVSRQNE
LDLILLASHW"
5042..33574
/gene="amphi"
5042..33574
/gene="amphi"
/notes="polyketide synthase multienzyme polypeptide housing
extension modules 9, 10, 11, 12, 13, and 14"
/codon_start=1
/transl_table=1
/product="Amphi"
/protein_id="AAK73501.1"
/db_xref="GI:14794893"
/translation="MNEOKRLDYLLATADLRARRRRVCELESASQEIATVGMTCR
YGGVSSPDLRMVZENGVTTFPTDRGMDLEALASFTSRGGFLHDAPEDEDADF
GISPREAVMDPQVRVLESWEAFERAGINTSVKSGRTGVFIGMAQDIYRGVADG
AGFQUTNGTSGVLSKRI SYTFGTGVPATVDTACSSSLVAHLTAQALRAEECTAL
AGGVYMSGRTGTEIEMRGQGLSVDRCSFGDADGTGWAEGVGLVLERISDAIRN
GREILAVVRGTAVNQDGSNGLTAPNGSQOAVIEQALVNARLSDGIDVVEAHGTGT
TLGDPVEAQAALLATYQOODEKPLLLGSVKSNISHTQAAAAGVAGVVKMVAWRHGV
PRTLADPETHRVNMQGAVRVLTENTEMPATCAPRAAVSFGISGTNAHTIVEQAP
EPEPADPDDPAAVTGVLVPLVLSGRSPVIRQAQAALLTLTGTPPPADUAY
SLATTATFEHRAVLASDLPELTGLRTAIAEGTDPAVLADTVTGTARTETRLAVLT
GQAQRLGARELAAFPVFMAALDAALDAFPLHDPVLRKVLWGSDADRLDRTETAYQ
PALFAVEVALYRLLESFVEKPDHLAGHSGEIAAAHVAGVLSDDAATLVAARGLMQ
ALPEGGAMVAQSEDEVAFLLAGHEDLSLAAVNGPSAVLSGSDTETTVTEAARLAA
DGRKTRLRVSHAFBPLMAPMLDFRNVEGLTLHSPLLPVVSDVTGEPATVAQLTS
PDVVDHVRQFAADGIDMLARHVDVTAFLGDPDSVLSMAQCLDQASDGLATVPA
LREGRPDTTAAALAAHQGTALHMDACFTGTGARTDLDLPTTQRRRYPRVAVQG
GAADLSVIGLGAHPHLLSAAVSLADSEGLLTGRI SLLSHMDADTVGRS VTYGRT
AFLEAVRAGDEVCGRVDELTAAPLVLPQEOGVQVQLWIGNPDASGRS VTYGRT
DADEADPWTSHANGVLSASRTTSDPDATVWPADAEPLVDGLYERLAEHGVGQPLF
QGLRAWRGDEVAEVLVPESGTDAESFGLHPALLDSALHNAHSAFVLDLRAAGSLP
FSMEGVS LHASGATLRLVRLAPAGDAVAIAVADDSQVLVLSADSLILRAAREIDA
AAALVRDALFRLDMPVPTVAASGTAAALVGEDPGLRALPQFGDLVAPHDADLAA
DGAAPDTVPLPLTGTPDADPVTAAHRAATEALAAVRLWEQDERFAASRLALVTGA
TTGHDPAARAAVGLVSAQSENPGRLVLDLADODTPALPAALSTSEEPOLAVRGE
LRARLVRPASTAEAVAPAGGEGAVLTGTTGGLGALVLAHLVAEHGVRLVSR
GMAAGAEVLVLAESGARATVACDVTDRAAVAEVLAHAPVSAVHSAAGVLDGCV
GLTLPRLTTLVRPKVDAAWNLHEATRDLDLKAFVLFSSVAGVLSGQANVAAAGNAF
LDALAARRAAGLPLGLSLAMGPWEQGTGTTGTTSEDDLRRMAPAGTTPALTVEQGLAL
DIALDGDAALAPVRLDLSVLRAGQVPPLLRLIRGRSRRRAAVAGSATAGLQAOLLA
RLDAGSRDELVLDRGOVALVGHATGAEI DAGRAPRELGPDSLTAVELRLNLTVT
GLRLPATLVDPYTVSHLASYLDLGTTEVAEVVQRTAAVADDPILVLMGACRYP
GGVTSBEDLRLVLTGTDVSGVPVNRGMDVENLYHPDPDPHGCTATRSGGFLHEAGE
FDPFGMGSPREALATDSQORLLLEASWEAREAGIDPVLGRSGTGVFAGVWYSDYS
AMLSGPFEGFGSGSPSLASGRVSYTLLEGPAVTVDTACSSSLVAHNAHQMLRS
GEISLALAGGVYVMSPTAVDFPARQGLSPGCRCKAFSDAGSDAGSGVGMQLVLER
QSDAIRNGHOLLVAVRGSANQDGSNGLTAPNGSQOQVIRQALASGGLSAGVDVW
EAHGTGTTGDPLEAQAALLATYGRDRDPEQPLLLGSVKSNIGHTQAAAAGVAGVKNM
SMRHGVLPRTLVHDVDASSHVDWTEGAVELLTEQTAMPETGCRPRRAVVSFGISGTNVH
TVLEQAPGTVFPARAAPTAGAPVLLLSGTRDRAQAARLLTLQNHHPSPSLADL
GHISLATTRTPPERAAVIAQDREGLLASLSLAAGRPDPAVVEGEAAHGAARVAFMSG
QGSORAMGRPELYTOPFAAADEVCALDPLDRLEVPFAAGSBEAALLDRTG
WTQPALFAVEVALYRLVESMGVADFTVTHSICEIAAAHTAGVTLODARLVAARAT
LMREALPSGMAMVAQATEEVALPLGGLSVAAVNGPSTSVVSGDEDDPAVELAAEFSG
RGKTRKLRVSHAFBSPHMDMLDAPFTVAETLSUAAAPRILVSDTLGTRADDAAEVRT
ADYVVRHREAVFADCVTRLDRAGATFLLELPGDGLLTAMAEITLDGDRYDHTALV
PLLRADRPBESAAATAAQIOTHQVLDLWDTAYLAGTCARRVOLPTFAFQHRWYWPQK
SAAPSAGPAPADQKAAVVERGDAELAAVLGDEDSLTPDLSLLPALSSWRNGQEK
ALLDTRVYVETRLSKPAPVLDGTTWLLVS DATADDETLDDGLAEALGAHGAIRV
RLVLDACDRAVRLGARLADTENADNTAQLVSLVLDERPTDGPAGFTOGLALTIYV
QALUADTGAGHRLWATRGAVSTGPADPTVTHPAQTAWGMGRGVALSHPRLMWGLVDLP
ADFDGAGQRLAEVLAVKQADGEDQVALRATGVGRRLVRIHVLDLSELPADQPTASGS
VLTITGTTGGTGAETARLARSAAHLVLTSSRRGPDAPGAELAELEAQSGASVSIAC

```

Alignment Scores:

Pred. No.:	0.000188	Length:	113193
Score:	654.50	Matches:	644
Percent Similarity:	30.30%	Conservative:	283
Best Local Similarity:	21.05%	Mismatches:	1086
Query Match:	4.95%	Indels:	1060
DB:	1	Gaps:	140

US-09-522-753-5 (1-2517) x AF357202 (1-113193)

QY	39	LeuLeuGluTyrGlnHisSerArgAspTyr-----AlaSerHis---52
DB	11581	CTCCTTCGCATCAGCGGCACCAATGTCACACCGCTCTGGACAGCGCGCGGCACAC 11640
QY	53	-----LeuSerProGlySerIleGln60
DB	11641	CGTCCG 11700
QY	61	ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln80
DB	11701	CGCACCG 11742

DVADRALAAVLDDGLSADQPLTGVVHTAGVGHYGLDALTAPAEFAGLTAAKLAGAAHL
 DNLLGDELDFFILFSGIAGVSGDSQAYGRANA YLDALALARRARRAGLAATIAWGP
 WGTGWAADAVSGTIRROGLGLDDPAPALTERRAVVRQDVTVTVADVDDWTYAPLF
 TARSALISDLPEVRALAEANTPADTGPASEIVQVRSLSEPEQLRLITDLVTRTAA
 TVLGHSSAGVADGAFREIGPDSLTAVELRKHGLCATGLSLPSTWFDVPTPLELAQ
 TYRAEWSVLEVADGAFREIGPDSLTAVELRKHGLCATGLSLPSTWFDVPTPLELAQ
 DPFVNRGWTAGLYDDPDHPGTTTSTGQFLHEADEFDPMFFGIGSPRALVNDMPQOR
 LLETTWEAFERAGLTPTDLRGLSTGTFITGSSYQYQYGMAGDGAEGHLVTGTPSVLS
 GRLAYVFLGEPAVTVTACSSSLVALHACQALRNGESNLAVAGCATVMTTNPANVA
 FSRORALADGRCKAFSESADGMTLAEGVGI VLVERLSDARRNNGHPLVAVIRGSAINO
 DGASNGLSAPNGSPQORVIRQALANARVAPGEIDLEAHGTGTPLGDTPEAOALFATY
 GRTRTPETALLGSKVNSIGHSQSAAAGVASI IKNWALRHGVMPQTLHADEPSSHVDW
 SPGTVRLGENTDMPTQTPRRAAVSSFGISGTNAHVILEQETEAFAADEQLAPAPL
 PVAAGVPMWLLSARGAALREQADRLTLHTVADPAARTIDIGLSLATRALFEHRAV
 VYPPAGTDDLEALRAVADGPGVARGVADVAGRTVTFVFPQSGWAGMAGQLLDES
 PVAFERIASCAAALAEFTDWNLDIVLRGAEGAPTLERVVDVOPASFAVMVSLAAWRA
 QGVBERDVAVGHSGQETAAAVSGLSLRGAERVTTRAQAI GRSLAGRGMMMSVALPV
 AEVEARLEAFEGRVSVAAENGRSSVAVAGEPEALDELHAQLTAETERRARVADYASH
 SHVEDLDEHLEILAEAVAPRTSEIFSTVTGDMLDTTVMQDGYWYRSLGRVLFAD
 AVRDLTIAADHRAPIEVSSHPVLAMSQMDIDDAGVAGVASGTLRRDNGGLDRFLLSAA
 EVFVRGVQVQWAAVFEGETGASRVOLPTYAFQHENLWMAAAPEAVTAADPEDAFWTA
 VEDGVSLTAALGTDEDSAAVLPALSSRRARKERTVDSWRYRTPKVPKLPQR
 TLDGTWLLVSADGVDDTVAEALETGGAERVLRLVLDSECTDRAVLERILTDDGLGTI
 VSLAGAERTIGAVPGTGLVGLVVALTVVALQALGDAGIDTPLWALTRGAVTSGRSKVT
 APVQOVTGIGMTALECPGRWGVVDLPETLDDARAGQALAAVLAGALGDDDDQIALRS
 SGVFTTIRI VRADAADPGSDARKWPGTTLTGGSGTLPALHARLAEQGAELHVLVSR
 RGPEPAAELAEALAEARTETTLAACTIDTDRDAVLAALLESKAEGRTVTVVHTAAT
 IELHTLDAITLDDFDRVLAAKVTGAOILDELDDDELDLDFVLYSSTAGMWSGAHAAY
 VAGNVAALAEHRRARGLTALSLSNGIWDADLOLGRVDPMQIRRSGLREMDPQALAS
 GUKAULDDQVIAVADWETIYHVITPPTLPDEVEVQRLTAARAEQSGADPAR
 GFFAALLAPAAEQHRLKLTETRTAAVLSGLSSAEDLTDQRAFDFGVGDSLUTAVGL
 RNRLASLGLTLPSTVTFVFPNPAALAGFLHSELADHAGAVAVTAGVAPDDDTAI
 VMGSCRYPGGITSAEOLMREVLSEVDAVSFPADRGDAEALYDDPDASGRYTVQG
 GLRDVADPDFGFGISPREALSMDPQORLLETLEWFEENAGLDDPVAQGRSTGTFI
 GASDYQGAAPVSGSESGHMITGSLPSVLSGRVSYLFGLEGPATVLTACSSSLVAI
 HLAQSLRNGESTLALAGASIMTSPMSFISQRLAEDNGRCKAYAGADQMTLAE
 GLGILLERLSDARRNGHEVLAVIRGSANQDGSNGLTAPNGSQOQVIRQALANAG
 VASANDIVLEHGHTGTALGDTPEAQALFATYGRDRDPEFVPLLSGYSKSNIGTHPMASSG
 FEANIIKLVHAREGVAPKSLHIDQPSRTHVDSGTTQLLTERTEWETGRPRRAVSS
 FGLSGTNVHTVLEQAADAPAEADTPADALVPVLVSGRGAALRAQAQALDILLLA
 ERPGHTDLSLATSRAALHRAAVADDHREALVRLGTLARDGLPGAGLVQGRTR
 GRTATLQGGSRGMRRELYRHPVFDALDVLARIDGTTTERPDLVLFPAEGSQ
 DAALLHRTGQAPALFALFALFLELSEVGTDPDLGHSVGEIAAAHVAGVVDLDDDE
 CTLVARGLMQALPEGGMVALAEAEDELPHLEGLAQDQSVAAVNGPRSVVADGEE
 EPLVLAHAFAGQRRTRKLRVSHAFADGICWLSLSTGGVQVTFLELPGDGLSWMARES
 TCTLAAPGVQCTADYVWVSHVSAVPAFCIGWLSLSTGGVQVTFLELPGDGLSWMARES
 LTDASKTALLPTLRGRPREQALVTAVAHAHGFVDVNTWMPQSGARRVALPTYAF
 QERYWPTTAAGITAPAGSALDAEFAAHEADVASLUTASLGLDDATVYAWVPA
 AWORRGEQSLDSWRYRVTWRPGGAGAAPTGRWLVLVPAEHRDEATAAWAADVEA
 ALATATVVEVTGTDRAALAEALTEAADGDTFGVLSLLALAPGDAGHGPAGPALTIT
 ATALQALGADARIDAPLWNI TRGAVAVGRSEQVTAPBQAAMWGLFRAAALLELPAVQGS

QY 81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
Db 11743 -----CACCCACTCCAGAACACACCCCGAGCGCTCCCT----- 11775
QY 101 GluPheileGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro 120
Db 11776 -----CGCCGACCTCGG-----CCACTCCCTGGCCACCCCG 11808
QY 121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysArgSer 140
Db 11809 CTCCTCGCTTCAGCGCGCGCGCG-----GGTGATCGCGCA 11844
QY 141 LeuThrGlyLys---LeuGluProValSerProSerProPro-----HisThrAsp 157
Db 11845 GG-ACCGGAGGAGTCTCGCTCCCTCGGCTCCCTCGCGCGCGCGCGCGCGCGCG 11903
QY 158 ProGluLeuGluLeuValProProArgLeu-----SerLys 169
Db 11904 CCGTCGTCGAGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11963
QY 170 -GluGluLeuileGlnAsnMetAspArgValAspArgGluileThrMetValGluGln 189
Db 11964 GCAGCCAGCG 12023
QY 189 nileSerLys-----Le 193
Db 12024 CGTTCGAGAGGTGTCG 12083
QY 193 uLysLysLysGlnGlnLeuGluGluAlaLysPro----- 207
Db 12084 TGTTCG 12143
QY 208 -----ProGluProGluLysProValSerProProL 219
Db 12144 CCGCGCTGTCG 12203
QY 219 eGluSerLysHisArgSerLeuValGlnleileTyrAspGluAsnArgLysLysAlaG 239
Db 12204 CCAGCTTCGTACCGGTCACTC-----CATCGGTGAGTCGCGCG 12242
QY 239 uAlaAlaHisArgileLeuGluGlyProGlnValGluLeuProLeuTyrAsnG 259
Db 12243 CCGCGCACATCGC-----CGGGGTGTTCACCTCCAGGAGCGCGCGCGCGCGCG 12296
QY 259 nProSerAspThrArgGlnTyrHisGluAsnleileLysleAsnGlnAlaMetArgLys 279
Db 12297 CCG-----CGCACCTCATGAGCGCGTGCCTG----- 12327
QY 279 sLeuileLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheC 299
Db 12328 -----CG 12360
QY 299 sGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgileGluAsnAs 319
Db 12361 -----GGAAG 12365
QY 319 nProArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProG 339
Db 12366 AGTCCG 12402
QY 339 uileArgLysGlnArg---GluLeuGlnGluArgMetGlnSerArgValGlyGlnArgG 358
Db 12403 ----CGCGGTCAACG 12458
QY 358 ySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluileleAspG 378
Db 12459 AGCTCCCGCGCGGTTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12516
QY 378 yLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValileProPhe 398
Db 12517 TGCCTTCCACTCG 12576

QY 398 tLeuTyrAspAlaAspGlnArgGlnArgileLysPheile---AsnMetAsnGlyLeuMetAl 417
Db 12577 GCTGTCTTACGCC---GCCCGCGCATCCCGTGTCTCCGACTTCCGACGGCGCGAGGGC 12633
QY 417 aAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlu 437
Db 12634 CGACGAGCGCGAGGTG-----CGCACCGCGACTACTGGTGGCGCCACGTCGG 12681
QY 437 sGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuileAlaSerPh 457
Db 12682 TGAGGCGGTACGG----- 12694
QY 457 eLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnG 477
Db 12695 -----TTGCGCGACTGCTGTCG-CACCTGCGCGACGCGCGCGCGCGCGCGCG 12737
QY 477 uAsnTyrLysSerLeuValArgSerTyrArgArgArgGlyLysSerGlnGlnGln 497
Db 12738 TCTTCTCGAACTCG 12797
QY 497 nGlnGlnGlnGlnGln----- 503
Db 12798 ACGAGCGTTACGACACACACACGCGACTCGTCCGCTGTCGCGCGCGCGCGCGCGCG 12857
QY 504 ----GlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLysAspG 522
Db 12858 AGAGCG 12917
QY 522 uLysGluLysGluLysGluAlaGluLysGluGluLysProGluValGluAsnAsp 542
Db 12918 CCGCGTACTGCG 12977
QY 542 sGluAspLeuLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysG 562
Db 12978 ACGCCCACTACTG----- 12990
QY 562 uAlaValAlaSer-----LysGlyArgLysThrAlaAsnSerGln----- 575
Db 12991 -CGCGAGTGCCTCCCG 13049
QY 576 -GlyArgArgLysGlyArgileThrArg-SerMetAlaAsnGluAlaAsn---SerGluG 594
Db 13050 GCGCGCGCGTGCAGCG 13109
QY 594 luAlaileThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerSera 614
Db 13110 ACAGCCTCACCCGCTGAGCTCCCTG-----CTGCCCGCGCTGAGCT 13151
QY 614 rGTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsn 634
Db 13152 CTTGCG 13211
QY 634 rPserAlaileAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrP 654
Db 13212 GACCGCGCTGAGCAAGCG 13242
QY 654 heAsnTyrLysLysArgGlnAsnLeuAspGluileLeuGlnGlnHisLysLeuLysMetG 674
Db 13243 -----CCTCGACG 13294
QY 674 luLysGlu-----ArgAsnAlaArgLysLysLysLysLysAlaProAlaA 689
Db 13295 ACCGAGCTGCTCGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13353
QY 689 laAlaSerGluGluAlaAlaPheProProValValGluAspGluMetGluAlaSerG 709
Db 13354 CTTCTGCTCC-----TGGACGCGCGACT----- 13373
QY 709 lyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyA 729
Db 13373 ----- 13373
QY 729 snGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrG 749

[illegible]

Db	14317	TGCCAAGCCCTACCTCGACGCGCCTCGCCTCGCGCCGCGCGCCCGGGCGCTCGCGCGAC	14377
Qy	1028	roProAlaAspLyysGluAlaPheAlaAlaGluAlaGlnLyysLeuProGlyAspProProc	1048
Db	14377	CTCCA-----TCGCCTGGGGTCCCTGGGGGGCACCGG	14409
Qy	1048	ysTrpThrSerGlyLeuProPheProValPro-----	1058
Db	14410	CATGGCGCGCGACGACCGCGTCTCGCGCACCTCGACGCCACAGGCGCTCGGCGTCTCGA	14469
Qy	1059	--ProArgGluValIleLyysAlaSerProHis-----	1069
Db	14470	CCCGCGC-----CGCCCTCACCAGATGCGCGCGCGCTGGTTCGCCGAGAGCT	14520
Qy	1069	laProAspProSerAlaPheSer-----TyrAlaProProGlyH	1082
Db	14521	CACGTCACCGTCGGGACGTCGACTGGACCGCTACGCCCGCTGTTTCACTTCGCGCG	14580
Qy	1082	isProLeuProLeuGlyLeuHis---AspThrAlaArgProValLeuProArg-----	1099
Db	14581	GCCCAGCGCCCTCATCAGCGACCTGCCCGAGGTCCGCTTGGCGCGCGAGAACACCC	14640
Qy	1099	roProThrIleSerAanProProProLeuIleSerSerAlaLyHisProSerValLeuG	1119
Db	14641	CGCGACACCGGCGACGCTCCGAG-ATCGTCACGGGTCCGCTCGCTGTCGAGACCGG	14699
Qy	1119	luArg-----luArg-----	1120
Db	14700	AGCACTGCGCTGCTCACCGACCTGGTGGCGCACCGCGCGACCGTGTCTCGTCACT	14759
Qy	1121	--GlnIleGlyAlaIleSerGln-----GlyMetSerValGlnL	1133
Db	14760	CCTCCGCGCGCGGTACCGGAGGACCGTCCCTTCGCGAGATCGCTTCGACTCGCTGA	14819
Qy	1133	euHisValProTyrSerGluHisAlaLyysAlaProValGly-----ProValThrM	1150
Db	14820	CGCCCTCGAACTGCGCAGACACCTCGGCGCGCCACCGCGCTCGTGTGCGACGACCA	14879
Qy	1150	etGlyLeuProLeuProMetAspProLyysLyysLeuAlaProPheSerGlyValLyysG	1170
Db	14880	T-GGTGTTGCACTACC---GACCCGCTGGAGCTGGCCCATGATCTCGCGGCCGAGATG	14935
Qy	1170	luGlnLeuSerProArgGlyGlnAlaGlyPro-----	1180
Db	14936	GTGCGCTCGTCTCGTAGGTGCG---GGGCCCCGCTCGCCACCGCGGTACCGACGACGA	14992
Qy	1180	-----	1180
Db	14993	CCATCGCATCATCGCANGAGTCGCGCTACCCGCGGAGTCAGCTCCCGGAGCAG	15052
Qy	1181	-----ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyT	1198
Db	15053	CTGTGGGACCTGGTGTCTTCG-GCACCGACCGGATCACCGACTTCC-----CGGTCAA	15105
Qy	1198	hrAlaLeuGlySerValProGly-----	1205
Db	15106	CCGCGCTGGAACACACCGCGGGCTCTACGACCCCGACCCGACCCCTGGCACCACTA	15165
Qy	1206	-----GlySerIleThrLyysGlyIleProSerThrArgValProSerAspS	1221
Db	15166	CTCCACGACGGCGGCTCTCTCGACGAGCGGACGAGTTTCGACCCCATGTTCTTCGGCAT	15225
Qy	1221	erAlaIleThrTyrArgGlySerIleThrHisGly-----	1232
Db	15226	CTGCCCCGGGAGCGCTGTCATGACCCGACGAGCGGCTCTCTCTGGAGACCACTG	15285
Qy	1233	-----ThrProAlaAspValLeuTyrLyysGlyThrIleThrA	1245
Db	15286	GGAGGCTTCGAACGCGCGGACTCACCCCG-GACACCTCC---CGCGCAGCCTGACCG	15341
Qy	1245	rgIleIleGlyLeuAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProL	1265
Db	15342	GCACCTTCATCGGCTCCAGCTACCGAGGATACGGATGTCGGCGCGGACGCGC---GCCG	15398

[illegible]

Qy	1541	euThrTyrGluAspHisGlyAlaPro-----PheAlaGlyHisLeuP	1555
Db	16397	CTCGTGACCGCCGACCGCGCGCCGCGATCGACATCGGTCTCTCCCTGGCCACCTCC	16456
Qy	1555	roArgGlySer-----ProValThrMetA	1563
Db	16457	CGCGCGC-TCCTCGAGCACCGCGCGTGTGTGCGCGCGCGACCGACCCGCTGGA	16515
Qy	1563	rgGluProThrProArgLeuGlnGluGlySerLeuSerSerSerlyysAlaSerGlnAspA	1583
Db	16516	GGCGCTCGCGCGCTCGCGCAGACGGGCGCTCGGCGCTCTCCCGCGCA	16565
Qy	1583	rgLysLeuThrSerThrProArgGluLeuAlaLysSerProHisSerThrValProGluH	1603
Db	16566	-----GTGTCGCGACGCTCGCGGCGCGCACCGTCTTCGTTCTCCCGCGCA	16611
Qy	1603	iHisProHisProLysSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspL	1623
Db	16612	GGGCTCCC-----ASTGGG-----	16625
Qy	1623	euTyrArgSerHisLeuProLeuAlaPhe--AspProThrSerIleProArgGlyIleP	1642
Db	16626	-----CGGGATGGGCGCCAGCTCTCTGACGAGTCCCGGTGTTCGCGAGCGATCGC	16680
Qy	1642	roLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThr-Tyr	1661
Db	16681	CG-----AATGCCCGCGCCCTCGCGGAGTTTCCGACTGGAA	16719
Qy	1662	ProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn	1681
Db	16720	CCTCATCGAGCTCTCGGGCGC-----CGAGGGCGCCCGAGCTGGAGCGCGTCGACGT	16776
Qy	1682	Arg-----GlnThrIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThr	1699
Db	16777	GTTCGACCGCGCTCGTTCGCGGTGATGTCTCTCTGGCGCGCTGTG-----CGGGGCCCA	16833
Qy	1700	AlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSer	1719
Db	16834	GGGTGTGCAACCGACGCGGTGTGCGGCACCTCGAGGCGCAGATCGCGCGCGGTGGT	16893
Qy	1720	LeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHis	1739
Db	16894	CTCGCGCGCTGTCTCTCGCGACCGCGCGCGGCTGTGACGTCTCCGCGCCAGGCCAT	16953
Qy	1740	LeuProValLeuValPro-----ProThrPro-----	1748
Db	16954	CGGCGCA-CCCTCGCGGGCGCGCGCATATGCTGGTGGCGTGCCTGCGCGGAGG	17012
Qy	1749	-----GlyThrProAlaThrAlaMetAspArgLeu	1758
Db	17013	TGGAGCGCGCTCGAGGCGTTTCGAGGACGGGTGTCTCGTCGCCGCCGAGA-----	17063
Qy	1759	AlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer-----	1773
Db	17064	-----ACGGCCCGCTCTCTCGCTGTCTCGCGCGCAGCCCGAGCGCCTGGACG	17111
Qy	1774	-----SerSerProLeuSerProGlyGlyPro-----Thr	1783
Db	17112	AATGCACGACAGCTCACCGCAGGAGATCGGGCCCGCGGTGCGCTGGACTACG	17171
Qy	1784	HisLeuThrLysPro-----ThrThrThrSerSerSerGluArgGluArg	1798
Db	17172	CCTCGCACTCCCCCATGTGAGGACCTGCACGACGAGATCTCGAACTGCTGGCCGAGG	17231
Qy	1799	AspArgAspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThr	1818
Db	17232	TCGCGC-----CGCGACGCTCGAGATCCCGTCTCTCTCGACCGCTCACCG	17276
Qy	1819	ThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSer	1838
Db	17277	GGCACTGGCTCGACACACCGTCATGAGCGCGGTACT-----	17315
Qy	1839	GlySerSerGlyGlyGlyGlyGlySerSerSerArg-----	1850

Db 17316 GGTACCGCAGCCTGCGCGCGCGTCTGTTTCGCGGACGAGTCCGTCGACCTGATCCCG 17375
Qy 1851 -----ProAlaSerHisSerHisAlaHisGlnHisSerProLysSerProArg 1866
Db 17376 CCGACACCGTGCCTTCATCGAGTCACTCCACCCGCGTCTCGCATGTCGCTCCAGG 17435
Qy 1867 ThrGlnAspAlaLeuGln-----ArgProSerValLeuHisAsnThrGly 1882
Db 17436 ACATGATCGACGACCGCGGAGTGGCGGTCTCGCTCCGACACCTCGCGCGGACACG 17495
Qy 1883 MetLysGlyLeuLeuThrAlaValGluProSerLys-----ProThrValLeuArgSer 1900
Db 17496 GCGGTCTCGACCGCTTCTCTGTCGCGCGCGGAGTGTTCGTCGCGGTTCAGGTG 17555
Qy 1901 ThrSer-----ThrSerSerProValArgProAlaAlaThrPheProProAlaThr 1917
Db 17556 ACTGGCGCGCGTTCGAGGGGACCGTGGTCCCGGTCGAGCTCCACCT-----17609
Qy 1918 HisCysProLeuGlyThrLeuAspGlyValThrProThrLeuMetGluProValLeu 1937
Db 17610 ---ATGCTTCCAGCACGACCACTGTGGCCATGGCGCGCGCGCGGAGGCG-17660
Qy 1938 LeuProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHis 1957
Db 17661 TCACCGCGCGGACCGCGAGGACCGCGCT-----17690
Qy 1958 AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPro-----AlaSer 1973
Db 17691 -----TCTGACCGCGCTCGAGCGGAGGACGTCTCCGCTCGACCGCGCGCTCG 17741
Qy 1974 SerProSerLysGlySerGluProArg-----ProLeuValProValSerGlyHis 1991
Db 17742 GCACGACGAGGACTCGTCCGCGCGTGTGCGCGCGTCTCTCTGCGCGGCGCC 17801
Qy 1992 AlaThrIleAla-----ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro 2009
Db 17802 GCAAGGAGCGCTCACCGTCACTCTCTGGCGCTACCGCGCGCTGGAAGCCGCTCACCA 17861
Qy 2010 AspProProAlaProProAlaSerAla-----SerAspProHisArgGluLys 2025
Db 17862 AGTCCCGGACGACCGCTCGAGCGGACCTGGTGTGTCGTCGCGCGGCGTCAAG 17921
Qy 2026 ThrGlnSerLysProPheSerIleGlnGluLeuArgSerLeuGlyThrHisGly 2045
Db 17922 ACACGAGCTGCGGAGCGCTGAGACGCGCGCGCGGAGTGGCG-17972
Qy 2046 SerSerThr-----2048
Db 17973 TCCTCTGGACGAGCTGCACCGACCGCGCGTGTGTCGCGGAGCGCTCACCGACGCG 18032
Qy 2049 -----SerProGluGlyValGluProValSerProValSerPro 2062
Db 18033 ACGGCTGACCGGATCGTCTCGTCTCGCGCGCGCGGAGCGGACCGGTGCGCTCCCG 18092
Qy 2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSer-----2080
Db 18093 GCACCGGCTCGTCTCGGCGTCCCTCACCGTCCGCGCTCGTCCAGGCGCTCGCGGACG 18152
Qy 2080 -----2080
Db 18153 CCGGTATCGACACCGCGCTGTGGCGCTGACCGCGCGCGCGCTCTCCACCGCGCGTCCG 18212
Qy 2081 ---HisLeuGluGlyGluLeuArgProLysGluPro-----GlyProValLys---2095
Db 18213 ACAAGGTACCGCGCGCTCCAGCGCGGACGTACCGGATCGGTGACCGCGCGCTG 18272
Qy 2096 -----LeuGlyGlyGluAlaAlaHisLeuPro---HisLeuArgProLeu---2109
Db 18273 AGTCCCGGACGCTGGGCGGTGTCGTCTGACCTGCGCGGAGACACTGGAGCGCGCGCG 18332
Qy 2110 -----ProGlu 2111
|||

Db 18333 GCACGCGGCTCGCGCGCGTCTCTCGCGCGCGCTCGCGGACGACGACGATCCCTCGC 18392
Qy 2112 SerGlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg 2131
Db 18393 GCTCTCGCGCGTCTTCCACCGCGTATCTCGCGCGCG-----18431
Qy 2132 ValValThr-LeuAlaGlnHisLeuGluValIleThrGlnAspTyrThrArgHisHi 2151
Db 18432 -----ACGCGCGCGCGCGTCTCGCGCGCGACTG-----GAAGCGCGCGCGCA 18479
Qy 2151 sProGlnGlnLeuSerAlaProLeuProAlaPro-----2162
Db 18480 CCCTGTGTACCG-GCGCTCTCGGACCTCGCCCGGACCTCGGCGCGCTGGCTCGCGAG 18538
Qy 2163 -----LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgAr 2178
Db 18539 CAGGCGCGCGAACACCTGTGTCTGTCAGCGCGCGCGCGCG-----AGCG 18586
Qy 2178 gProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPr 2198
Db 18587 CCGCGCGCGCGCACTCCGCGCGGAGTGGCGGACGCGGCGACCGAGACGCTCGCC 18646
Qy 2198 oHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerVal-LeuGlyGly 2218
Db 18647 GCTGTGACATCACCGACCGTGTGCGCGCTCGCGCGCTCTCGAAAGCCTGAAGCGCGAG 18706
Qy 2218 LyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerA 2238
Db 18707 GCGCG-----CACGCTCGGACGCGTGTGTGCACCGCGCGCGCACCATC 18748
Qy 2238 rgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetG 2258
Db 18749 GAACCTGCA-----CACCTCGACGCGCA-C 18771
Qy 2258 LySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluS 2278
Db 18772 CACCTCGACGACTTCGACCGCGCTCTCGCGCGCAAGGTCAACCGCGCGGATCTCTGA 18831
Qy 2278 erAsnSerAlaMetValLysSerLysGlnGluLeuAsnLysLeuAsnThrHisA 2298
Db 18832 CGAATCTCTCGACGAGGAGC-----TGGACGACTTCGT 18867
Qy 2298 snArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProA 2318
Db 18868 CCGTACTCTCTCA-----CCG 18885
Qy 2318 laIleThrGlyThrGlyLeuMetThrTyrArg---SerGlnAlaValGlnGluHisAlas 2337
Db 18886 CGCATGTGGGCGAGCGCGCGCGCGCGCGTATGTGCGCGGCAACG-----18932
Qy 2337 erThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTyrAspGlnT 2357
Db 18933 -----CCTATCT 18939
Qy 2357 rpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerL 2377
Db 18940 GCGCGCGCTCGCG-----AGCACCGCGCG 18963
Qy 2377 euProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerP 2397
Db 18964 CGCGCGCGCGCTGACCGCGCTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 19023
Qy 2397 rogly-----GlyGlyGlyLysAlaLysValSerGlyArgProSerS 2411
Db 19024 CGCGCGAGTGCACCGCGAGTATGTCGCGCGCGCGCTGGAGTTCATGACCGCGAGCT 19083
Qy 2411 erArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValS 2431
Db 19084 CGCGCTGAGCGCGCTCAAGCGCGCGCTCGACGAGC-----AGCAGGT 19128
Qy 2431 erSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpG 2451
Db 19129 CATCGCGGTGCGCGAGCTCGACTCGGAGAGCTACACCGCGCTCTACACCTCGCGCGCGCC 19188
|||


```
Qy 457 eLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnG1 477
Db 12695 -----TTGCGCGACTGCTCCG-CACCGTCGGCGAGCGCGGCCACCC 12737
Qy 477 uAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnG1 497
Db 12738 TCTTCTCGAACTCGGCGCGCGAGCTGTGACCGCATGCGCGAGACACCTCGCGCG 12797
Qy 497 nGlnGlnGlnGlnGln-----503
Db 12798 ACGAGCGTTACGACCAACACACGCGCACTGCTCCCGCTGCTGGCGCGCGCGCCGAGG 12857
Qy 504 -----GlnGlnGlnGlnGlnProMetProArgSerSerGlnGluLysAspG1 522
Db 12858 AGAGCGCGCGCGCGCGCGCGCGCTCCAGATCCACCGCGTCCGACTCGACTGGA 12917
Qy 522 uLysGluLysGluLysGluAlaGluLysGluLysGluLysProLysValGluAsnAspLy 542
Db 12918 CCGGTTACTGCGCGCGCGCGCGCGCGCTGCGACCTGCGCGACTGCGCTTCGAGC 12977
Qy 542 sGluAspLeuLysGluLysThrAspThrSerGlyGluAspAsnAspGluLysG1 562
Db 12978 ACGCCCACTACTG-----12990
Qy 562 uAlaValAlaSer-----LysGlyArgLysThrAlaAsnSerGln-----575
Db 12991 -GCCGAGCTGCCCTCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGAGCTGT 13049
Qy 576 -GlyArgArgLysGlyArgLysThrArg-SerMetAlaAsnGluAlaAsn---SerGluG 594
Db 13050 GGCGCGCGCGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13109
Qy 594 luAlaIleThrProGlnGlnSerAlaGluLeuAlaSerMetGluLeuAsnGluSerA 614
Db 13110 ACAGCTCACCGCGCTGACTCCCTG-----CTCGCGCGCTGAGCT 13151
Qy 614 rgTrpThrGluGluGluMetGluThrAlaLysLysGlyLeuLeuGluHisGlyArgAsnT 634
Db 13152 CTTGGCGCGCGCAACAGGAGAGGCGCTCTCGAGACGCTGCGCTACCGGCTCGAGT 13211
Qy 634 rpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPhetYrP 654
Db 13212 GSACCGCGCTGAGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13242
Qy 654 heAsnTyrLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetG 674
Db 13243 -----CTCGAGCGCACTGCTGCTGCTGCTCTCTCTCGAGCGCGCGCGCGCGCGAG 13294
Qy 674 lylusGlu-----ArgAsnAlaArgArgLysLysLysLysAlaProAlaA 689
Db 13295 ACCGAGCTGCTCGAGCGCGCTGGCGAGGCGCTGGCGCGCGCGCGCGCGCGCGCGCG 13353
Qy 689 laAlaSerGluGluAlaAlaPheProProValValGluAspGluMetGluAlaSerG 709
Db 13354 CTTGCTGTC-----TGAGCGCGCGACT-----13373
Qy 709 lyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAlaSerGlyA 729
Db 13373 -----13373
Qy 729 snGluValProArgGlyGluCysSerGlyProAlaThrValAsnAsnSerSerAspThrG 749
Db 13374 -----GGCGCGCGCGCGCGCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCTG 13406
Qy 749 luSerIleProSerProHisThrGluAlaAlaLysAspThrGlyGlnAsnGlyProLysP 769
Db 13407 -----CCGACACCGAGAACCGCGCAACA-----C 13431
Qy 769 roProAlaThrLeuGlyAlaAspGlyProProGlyProProGlyProProThrProArgArgT 789
Db 13432 CGCCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13491
Qy 789 hrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProP 809
```

```
Db 13492 CACCCAGGGCTGCGCTGACCATCGCTCGTCCAG-----CCCTCGCGCA 13539
Qy 809 roProAlaProProSerProSerAlaProProValValProLysGluLysGluG 829
Db 13540 CACCGCGCGCGCGCGCTGTGAGACCGCACCGCGCGCGCTCTCCACCGACCGCG 13599
Qy 829 luGluThrAlaAlaAlaProProValGluGluGlyGluGluLysProProAlaAlaG 849
Db 13600 CGACCGGTCACCCACCGCGCG-----AGCCACCGCTGGGG 13638
Qy 849 luGluLeuAlaValAsp-----ThrGlyLysAlaGluG 860
Db 13639 CATGGCGCGCGTGTCCCTCGGAGCACCCCGCGCTGTGGCGCGCTCGTGCACCTCCC 13698
Qy 860 luProValLysSerGluCysThrclucIuAlaGluGluGlyProAlaLysGlyAspA 880
Db 13699 GGCGGACTTCAGCGCGCGCGCGCGCGCTCGCGAGGTCTCTCGGTCGAGGACGC 13758
Qy 880 laGluAlaAlaGluAlaThrAlaGluLysAlaGluLysGluLysGlyGlyS 900
Db 13759 CCGCGCGCGAGGACGAGTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCTCGT 13818
Qy 900 erGlyArgAlaThrAla-----LysSerSerGlyAlaProGlnAspSerA 916
Db 13819 CCGCCACATCGTCGAGCTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 13878
Qy 916 spSerSerAlaThrCysSerAlaAspLysValAlaGluGlyGlyAspLysAsnA 936
Db 13879 GATCACCGCGCGCGCGCGCGCGCG-----13904
Qy 936 rgLeuLeuSerProArg-----ProSerLeuLeuThrProThrGlyAspP 951
Db 13905 -----GGCGCGAGACCGCGCGCTGCTGGCGCGCGCGCGCGCGCGCGCT 13956
Qy 951 roArgAlaAsnAlaSerProGlnLysProLeu-----AspLeuLysG 965
Db 13957 CACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCA 14016
Qy 965 lnLeuLysGlnArgAlaAlaAlaIleProProIleGlnValThrLysValHisGluProP 985
Db 14017 GTCGGGCGCAAGCGTCTCGATCTGCGCTCGCGCTGCGCGCGCGCGCGCGCGCT 14076
Qy 985 roArgGlu-----AspAlaAlaProThrLysProAlaPro-----996
Db 14077 GTCTCTCGAGCGACTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14136
Qy 997 -----ProAlaProProProp 1002
Db 14137 CGGCCACTACGGCGCGCTGAGCGCGCTGAGCGCGCGCGCGCGCGCGCGCTCACCGCGC 14196
Qy 1002 roGln-----AsnLeuGlnProG 1008
Db 14197 CAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 14256
Qy 1008 luSerAspAlaProGlnGlnProGlySerSer-----1018
Db 14257 CATCTGTTGCGTCCATCGCGCGCGCTCTGGGCGAGCGCGCGCGCGCGCGCT 14316
Qy 1019 -----ProArgGlyLysSerArgSerProAlaP 1028
Db 14317 TGCCAAAGCGCTACTCTGAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 14376
Qy 1028 roProAlaAspLysGluAlaPheAlaAlaGluAlaGlnLysLeuProGlyAspProP 1048
Db 14377 CTCCA-----TCGCTGGGCTCTCTGGGCGCGCGCGCGCGCGCGCGCGCG 14409
Qy 1048 ysTrpThrSerGlyLeuProPheProValPro-----1058
Db 14410 CATGGCGCGCGAGCGCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCT 14469
Qy 1059 --ProArgGluValIleLysAlaSerProHis-----A 1069
```

Db 14470 CCCCCGCGC-----CCGCCCTCACGAGATGCGCGCGCGTGGTCCGCCAGGACGT 14520
Qy 1069 laProaspProSerAlaPheSer-----TyAlaPProGlyH 1082
Db 14521 CACCGTACCGTCCGGAGCGTCACTGGACCGCGCTAGCGCGCGCTGTTTCACTCCGCCG 14580
Qy 1082 isProLeuProLeuGlyLeuHis---AspThrAlaArgProValLeuProArg-----P 1099
Db 14581 GCCCGCGCCCTCATCAGCGACCTGCCGAGGTCCGTCGCTGCCGCGCGAGAACACCC 14640
Qy 1099 roProThrIleSerAsnProProLeuIleSerSerAlaIysHisProSerValLeuG 1119
Db 14641 CGCCGACCGCGGCGCGCTCCGAG-ATCGTCCAGCGGGTCCGCTCGCTGTCGGAGCCG 14699
Qy 1119 luArg----- 1120
Db 14700 AGCACTGCGCGCTGTCAACGACTGGTGGCAACGAGGCGCGGACCGTGTCTGCTCACT 14759
Qy 1121 --GlnIleGlyAlaIleSerGln-----GlyMetSerValGlnL 1133
Db 14760 CTTCCCGCGCGCGTACCGGAGGACCGTGCCTTCCGCGAGATCGGCTTCGACTCGCTGA 14819
Qy 1133 euHisValProTyrSerGluHisAlaIysAlaProValGly-----ProValThrM 1150
Db 14820 CCGCGCTCGAACTGCGCAAGCACCTCGCGCGCGCCACCGGCTCTCGCTGCCGAGCACCA 14879
Qy 1150 etGlyLeuProLeuProMetAspProIlyAlaIysLeuAlaProPheSerGlyValLysGlnG 1170
Db 14880 T-GGTGTTCGACTACCC--GACCCGCTGGAGCTGGCCCGACGATCTCGGGCCGAGATG 14935
Qy 1170 luGlnLeuSerProArgGlyGlnAlaGlyPro----- 1180
Db 14936 GTCGGCTCGTCTCGAGGTGCG--GGGCCCCGTGCGCACCGCGGTACCGACGACGAA 14992
Qy 1180 ----- 1180
Db 14993 CCCATGCCATCATCGGCATGAGTGCCTACCCCGCGGGAGTCACTCCCGCGAGCAG 15052
Qy 1181 -----ProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyT 1198
Db 15053 CTGTGGACCTGTGTCTCCG-GCACCGAGCGGATCACCGATTCC-----CGGTCAA 15105
Qy 1198 hrAlaLeuGlySerValProGly----- 1205
Db 15106 CCGCGGTGGAACACCGCGCGCTTACGACCCCGCGACCCCGACCCCTGGCCACCACTTA 15165
Qy 1206 -----GlySerIleThrLysGlyIleProSerThrArgValProSerAspS 1221
Db 15166 CTCACGCGAGGCGGCTTCTGACGAGGCGGACGAGTTCGACCCCATGTCTTCGGCAT 15225
Qy 1221 erAlaIleThrTyrArgGlySerIleThrHisGly----- 1232
Db 15226 CTGCGCGCGAGCGGTGTGTATGACCCGCGAGCGGCTCTCTGGAGACCACTG 15285
Qy 1233 -----ThrProAlaAspValLeuTyrLysGlyThrIleThrA 1245
Db 15286 GGAGGCTCTGAACGCGCGGACTCACCCCG-GACACCTC-----CGCGCGAGCTGACCG 15341
Qy 1245 rgIlelleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProL 1265
Db 15342 GCACCTTCATCGGCTCCAGCTACCGAGGTACGGGATGGGCGCGCGCGCGC-----GCCG 15398
Qy 1265 ysGlyHisVal-----IleTyrGluGlyLysLysGlyHisVal 1278
Db 15399 AGGGCACCTGTGTACCGGACACGACCCGAGCGGTCTCTCGCGCGGCTCGCTATGTCT 15458
Qy 1278 euSerTyrGluGly---GlyMetSerVal---ThrGlnCysSerLysGluAspGlyArgS 1296
Db 15459 TCGGCTTGGAGGCGCGCGGTACCGTTCGACACCGGCTGCTGCTCTCGCT--GGTCGCC 15517
Qy 1296 erSerSerGlyProProHisGluThrAlaIaProLysArgThrTyrAspMetMetGluG 1316
Db 15518 CTGACCTGCGCTGCCA-----GGCGCTCGCAACCGCGCGGCTCGCTCGCGGTGCGG 15571

Qy 1316 lyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleP 1336
Db 15572 GCGCGCGCACCGTATGATCACCAACCCCAACGCGTTCGTGGCTTACGCGCGGCGCGCC 15631
Qy 1336 ro-----ProGluArgHisSerProHisH 1344
Db 15632 CTCGCGCAGCAGCGCGGTGCAAGCGTTCTCCGAGAGCGCGCATGACCTTCGCC 15691
Qy 1344 isLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrV 1364
Db 15692 GA---AGGTGTCCGATCTCTCTGTCGAGCGGCTGTCGACCGCGCGCGCAACGCGCAC 15748
Qy 1364 alGluAlaGlnAsp---TyrLeuArgArgGlu---AlaLysLeuLeuLysArgGluG 1382
Db 15749 CCGGTCTCGCGTGTATCGCGGCTCCGCTATCAACACGAGCGCGGCTCCCAACGCGCTC 15808
Qy 1382 lyThrPro-----ProProProPro----- 1388
Db 15809 AGCGCCCCCAACGCGCCCTCCGACGAGCGGCTATCCGCGAGCGCTTGCCCAACGCGCG 15868
Qy 1389 -----ProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysL 1407
Db 15869 GTGCGCCCCCGGCGAGATCGACTCTCTGA-----GGCCCCACGCGACC 15910
Qy 1407 euLysProAla-----HisGluGlyLeuValAlaThrValLysGluAlaGly--- 1422
Db 15911 GGCACCCGCTCGGCGACCCCATCGAGC-----GCAGGCGCTG 15949
Qy 1423 --ArgSerIle-----HisGluIleProArgGlu-----GluL 1433
Db 15950 TTGCGCACCTACGCGCGCACCCGTACCCCGCGAGACCGCGCTGCTCGCTCGGTGAAG 16009
Qy 1433 euArgHisThrProGluLeuProLeu----- 1441
Db 16010 TGAACATCGG--GCACCTCCAGTCCGCGCGGTGTCGCGAGCATCATAGATGGTC 16066
Qy 1442 -----AlaProArgProLeuLysGluGlySerIleThrGlnG 1454
Db 16067 ATGGCGCTCGCCACGCGTGTATGCCCGACGCTGACGCGCGAGCGCTC----- 16119
Qy 1454 lyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysHisAspVal- 1473
Db 16120 --CAGCCACGTGACTGGAGCGCGGCGACCTACGCGTCTCGCGGAGAACACCGACTGG 16177
Qy 1474 -----ArgSerLeuIleGlySerProGlyArgThrPheP 1485
Db 16178 CCGCAGCGGCGCTCGCGCGCGCGGTGCTCTCTCGGCATCAGCGTACCAAC 16237
Qy 1485 roProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrG 1505
Db 16238 GCCACGTCTCT-----GGAGCAGGAGACCGCGCGCGCC 16276
Qy 1505 luGluSerLeuLysSerArgProGlyThrAlaSerSerGly-----GlySerIleA 1523
Db 16277 GCGAGGACGAGCAGCTCGCGCGCGCGCTCCCGCGTCCCGCGCGCGGTGTGGTGGCGTG 16336
Qy 1523 laArgGlyAlaPro-----ValIleValProGluLeuGlyLysProArgGlnSerProL 1541
Db 16337 CTGCTCTCGCGCTGGCGCGCGGTGCGCGAGCAGCGCGCTCTCTCACCCTAT 16396
Qy 1541 euThrTyrGluAspHisGlyAlaPro-----PheAlaGlyHisLeuP 1555
Db 16397 CTGTCACCGCGCACCGCGCGCGCTCGACATCGGTCTGCTCTCTCGCGCACCTCC 16456
Qy 1555 roArgGlySer-----ProValThrMeta 1563
Db 16457 CCGCGCGC-TCTTCGACACCGCGCGCGGTGTGTCGCGCGCGCGCGCACCCGCGCTGGA 15515
Qy 1563 rgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGlnAspA 1583
Db 15516 GCGCTGCGCGCGTCCGCGCAGACCGGCGCTCGCGCGCG----- 15655

Qy 1583 rglyLeuThrSerThrProArgGluLeuAlaIysSerProHisSerThrValProGluH 1603
Db 16566 -----GTGTCGGAGACGTGGGGCGGCACCGTCTCTCTCCCGGCCA 16611
Qy 1603 iHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspL 1623
Db 16612 GGGCTCCC-----AGTGGG----- 16625
Qy 1623 euTyrArgSerHisIleProLeuAlaPhe---AspProThrSerIleProArgGlyIleP 1642
Db 16626 -----CGGGATGGGCCAGCTCTCTGGAGAGTCCCGGTGTTCGCCAGCGATCGC 16680
Qy 1642 roLeuAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThr-Tyr 1661
Db 16681 CG-----AATCGCGCGCGCTCGCGAGTTCCACGACTGAA 16719
Qy 1662 ProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsn 1681
Db 16720 CCTCATCGACGTCTCGCGGGCGC---CGAGGGCGCCCCGACGCTGGAGCGCGTTCGACGT 16776
Qy 1682 Arg-----GlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisAsnThr 1699
Db 16777 CGTCCAGCCGCTCTGTCGGGTGATGTCTCCTCGCGCGGTGTG---GGGGGCCA 16833
Qy 1700 AlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSer 1719
Db 16834 GGGTGTGAAACCGACGCGGTGGTGGCCACTCGACGGCGAGATCGCCGCGGTGTG 16893
Qy 1720 LeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHis 1739
Db 16894 CTCGGCGCGCTGTGCTGCGAGCGCGCGGTCTGTGACGCTCGCGCCCCAGGCCAT 16953
Qy 1740 LeuProValLeuValPro-----ProThrPro--- 1748
Db 16954 CGGCCGCA-CCCTCGCGCGCGCGCGCATGATGTGCGTGGCGCTCCCGTCCCGAGG 17012
Qy 1749 -----GlyThrProAlaThrAlaMetAspArgLeu 1758
Db 17013 TGGAGGCGCGTCTGGAGCGGTTCGAGGACGGGTGTCGTCGCGCGCGAGA----- 17063
Qy 1759 AlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSer----- 1773
Db 17064 -----ACGGCCCCGCTCTCTCGTTCGCGGGCGAGCGCCGAGCGCCCTGGACG 17111
Qy 1774 -----SerSerProLeuSerProGlyGlyPro-----Thr 1783
Db 17112 AACTGACGACAGCTCACCGCGAGAGATCCGGGCCCGCGGTGCGCGTGCATGAC 17171
Qy 1784 HisLeuThrLysPro-----ThrThrThrSerSerSerGluArg 1798
Db 17172 CCTCGCACTCCCCCATGTCGAGGACCTGCACGACGAGATCTCTGAACTGTCGCGCAGG 17231
Qy 1799 AspArgAspArgGluArgAspArgAspArgGluArgGluIysSerIleLeuThrSerThr 1818
Db 17232 TCGCGC-----CGGGACGTCCGAGATCCCGTCTCTCGACCGTCAACG 17276
Qy 1819 ThrThrValGluHisAlaProIleTyrArgProGlyThrGluGlnSerSerGlySerSer 1838
Db 17277 GCGACTGGCTGCACACCGCTCATGAGCGCGGTACT----- 17315
Qy 1839 GlySerSerGlyGlyGlyGlySerSerSerArg----- 1850
Db 17316 GGTACCGCAGCTCGCGCGCGGTGTGTTGGGGAGCGAGTCCGTGACCTGATCGCG 17375
Qy 1851 -----ProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg 1866
Db 17376 CCGACCAACCGTGCCTCATCGAGTCACTCCACCCCGGTGCTGCCATGTCCGTCCAG 17435
Qy 1867 ThrGlnAspAlaLeuGlnGln-----ArgProSerValLeuHisAsnThrGly 1882
Db 17436 ACATGATCGAGCGCGGAGTGGCGGTGTCGCTCCGCGACCCCTCGCGCGGACACG 17495
Qy 1883 MetLysGlyIleIleThrAlaValGluProSerLys-----ProThrValLeuArgSer 1900

Db 17496 GCGGTCTCGACCGCTTCTCTGTCGCGCGCGAGGTGTTCGTCCGCGGTGTCAGGTG 17555
Qy 1901 ThrSer-----ThrSerSerProValArgProAlaAlaThrPheProAlaThr 1917
Db 17556 ACTGGCGCGCGGTGTCAGGAGACCGTCTGCTCCCGGTGAGACCTCCACCT----- 17609
Qy 1918 HisCysProLeuGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeu 1937
Db 17610 ---ATGCTTCCAGCAGAACTGTGGGCCATGGCGCGCGCGCGAGGCG----- 17660
Qy 1938 LeuProLysGluAlaProArgValAlaAlaProGluArgProArgAlaAspThrGlyHis 1957
Db 17661 TCACCGCGCGCGCGCGCGCGCGT----- 17690
Qy 1958 AlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPro-----AlaSer 1973
Db 17691 -----TCTGACCGCGTGCAGACGGGACGTCTCCGCCCTGACCGCGCGCTCG 17741
Qy 1974 SerProSerLysGlySerGluProArg-----ProLeuValProProValSerGlyHis 1991
Db 17742 GCACCGACGAGACTCGTCCGCGCGGTGTGTCGCGGTCTCTCTGCGCGCGGCC 17801
Qy 1992 AlaThrIleAla-----ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerPro 2009
Db 17802 GCAAGGAGCGCTCCACCGTGCAGTCTCTGGCGCTACCGCGCGACCTGGAAGCCCTCACA 17861
Qy 2010 AspProAlaProProAlaSerAla-----SerAspProHisArgGluLys 2025
Db 17862 AGTCCCGCAGCGACCTCGACGGACCTGGGTGTGTGTCGCGCGCGGTGTCGACG 17921
Qy 2026 ThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGly 2045
Db 17922 ACACCGCGTCCGCGAGCGCTGAGACGGCGCGCGGTGCGCC-----GGC 17972
Qy 2046 SerSerTyr----- 2048
Db 17973 TCCTCTGGACGAGAGTGCACCGACCGCGCGGTGTGTCGCGAGCGCTCACCAGCGCG 18032
Qy 2049 -----SerProGluGlyValGluProValSerProValSerSerPro 2062
Db 18033 ACGGCTGACCGCGATGCTCTCGTCTCGCGGGCGCGAGCGGACCGGTGCGCTCCCG 18092
Qy 2063 SerLeuThrHisAspLysGlyLeuProLysHisLeuGluLeuAspLysSer----- 2080
Db 18093 GCACCGGTCTCTCTCGGGTCCCTCACCGTCCGCTCGTCCAGCGCTCGCGCAGC 18152
Qy 2080 ----- 2080
Db 18153 CCGGTATCGACACCGCGCTGTGGCCCTGACCGCGCGCGCTCTCCACCGCGCTCCG 18212
Qy 2081 ---HisLeuGluGlyGluLeuArgProLysGlnPro-----GlyProValLys--- 2095
Db 18213 ACAAGGTACCGCGCGCTCCAGCGCGAGTCCAGGATCGGTGAGACCGCGCGTGG 18272
Qy 2096 -----LeuGlyGlyGluAlaAlaHisLeuPro---HisLeuArgProLeu--- 2109
Db 18273 AGTCCCGGACCGTGGGGCGGTGTGTCGACCTGCGCCGAGACACTGAGCGCGCGCG 18332
Qy 2110 -----ProGlu 2111
Db 18333 GCCAGCGGTCCGCGCGCTCTCGCGCGCGGTGCGCGAGCAGCAGATCGCCTGC 18392
Qy 2112 SerGlnProSerSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArg 2131
Db 18393 GCTCTCCGCGCGTCTTACCCCGGTATCTCGCGCGCG----- 18431
Qy 2132 ValValThr-LeuAlaGlnHisIleSerGluValIleThrGlnAspTyrThrArgHisHi 2151
Db 18432 -----ACGCGCGCGCGCGGTCCGCGCGACTG-----GAACCGCGCGGACCA 18479
Qy 2151 sProGlnGlnLeuSerAlaProLeuProAlaPro----- 2162

```
Db 18480 CCCTGCTCACCG -GCGGCTCGGCGACCCCTCGCCCCGACCTCGCGCGCTGGCTCGCCGAG 18538
Qy 2163 -----LeuTyr-SerPheProGlyAlaSerCysProValLeuAspLeuArgAr 2178
Db 18539 CAGGGCGCCGAACACCTGCTCTGCTGTCAGCCGCGCGGCCGCG-----AGGCG 18586
Qy 2178 gProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerPr 2198
Db 18587 CCCGCGCGCCGAACCTCGCGCGCGAGCTGGCGGAACGCGGACCGAGACGACCTCGCC 18646
Qy 2198 oHisSerGluGlyGlyArgSerProGluProAenLysThrSerVal-LeuGlyGlyG 2218
Db 18647 GCCTGCGACATCACCGACCGCTGACCGCTCGCGCGCTCTCGAAGCCTGAAGCGCGAG 18706
Qy 2218 lylGluAspGlylleGluProValSerProProGluGlyMetThrGluProGlyHisSerA 2238
Db 18707 GCGCG-----CACCGTCCGACCGGTGGTGACACCGCGCGCCACCATC 18748
Qy 2238 rgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetG 2258
Db 18749 GAACCTGCA-----CACCTCGAGGCCA-C 18771
Qy 2258 lySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluS 2278
Db 18772 CACCCTCGACGACTTCGACCGCTCTCGCGCGCAAGGTCAACGCGCGCGAGATCTCGA 18831
Qy 2278 erAenSerAlaMetValLysSerLysGlnGluLeuAenLysLysLeuAenThrHisA 2298
Db 18832 CGAACCTCTCGACGACGAGAGC-----TGACGACACTCGT 18867
Qy 2298 snArgAenGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProA 2318
Db 18868 CCTGTACTCTCCA-----CCGC 18885
Qy 2318 laIleThrGlyThrGlyLeuMetThrTyrArg---SerGlnAlaValGlnGluHisAlaS 2337
Db 18886 CGGCATCTGGCGGACGCGCGCACCGCGGTATGTGCGCGGCAACG----- 18932
Qy 2337 erThrAenMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnT 2357
Db 18933 -----CCTATCT 18939
Qy 2357 rpGluGluSerProProLeuSerAlaAlaPheAsnProLeuAsnAlaSerAlaSerL 2377
Db 18940 GCGCGCCCTCGCG-----AGCACCGCGC 18963
Qy 2377 euProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerP 2397
Db 18964 CGCCCGCGGCTGACCGCCCTGCTGCTGCTGCGGCATCTGCGCGCACGACCTCCAGCT 19023
Qy 2397 roGly-----GlyGlyGlyLysAlaLysValSerGlyArgProSerS 2411
Db 19024 CGGCCGAGTGCACCCGAGATGATCCCGCAGCGGCTCGAGTTCATGACCGCGAGCT 19083
Qy 2411 erArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSerValS 2431
Db 19084 CGCGCTGAGCGGCTCAAGCGGCGCTCGACGACG-----AGCAGGT 19128
Qy 2431 erSerValHisSerGluGlyAspCysAenArgArgThrProLeuThrAsnArgValTrpG 2451
Db 19129 CATCGCGTCCGCGACGCTCGACTGGGAGAGTACACCCCGCTCTACACCTCGCGCGCGCC 19188
Qy 2451 luAspArg-----ProSerSerAlaGlySerThrProPheProTyrAsnP 2466
Db 19189 CACCCGCTGTTCCAGCAGGTCCTCCGAGGTCCAGCGCTACCCGCGCTCGCGAGCAGA- 19247
Qy 2466 roLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProProGlyLeuProA 2486
Db 19248 -----GCGCGCGCGGACCCCGC 19263
Qy 2486 laGlySerGlyProLeuAlaGlyProHisAlaTTP----- 2498
Db 19264 CCGGCGCG-----AGTTCGCGCGCGCGCTGCTGCGCCCTGCGCCCGCGCCGACGACA 19314
```

```
Qy 2499 -----AspGluGluProLysProLeuLeuCysSer 2508
Db 19315 CGCAAGCTGCTGGAGACCGCTCGCACCGAGCGCGCTCGTCTCG 19361

RESULT 98
SPSNBCDE 22449 bp DNA linear BCT 25-APR-1997
LOCUS S.pristinaespiralis snbc gene & snbDE gene.
DEFINITION Y11548
ACCESSION Y11548.1 GI:2052247
VERSION Pristinamycin I synthase 2; Pristinamycin I synthase 3 and 4; snbC
KEYWORDS gene; snbDE gene.
SOURCE Streptomyces pristinaespiralis
ORGANISM Streptomyces pristinaespiralis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1
AUTHORS de Crecy-Lagard, V.A., Saurin, W., Thibaut, D., Gil, P., Naudin, L.,
Crouzet, J. and Blanc, V.
TITLE Streptogramin B biosynthesis in Streptomyces pristinaespiralis and
Streptomyces virginiae: molecular characterization of the last
structural peptide synthase gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 22449)
AUTHORS de Crecy-Lagard, V.A.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1997) V.A. de Crecy-Lagard, Institut Pasteur,
Unit de Biochimie Microbienne, 28 rue du Dr Roux, 75724 Paris
cedex, FRANCE

FEATURES
source
1..22449
/organism="Streptomyces pristinaespiralis"
/mol_type="genomic DNA"
/strain="SP92"
/db_xref="taxon:38300"
/clone_lib="PIBV3 cosmid"
115..7890
/gene="snbc"
115..7890
/gene="snbc"
/codon_start=1
/transl_table=11
/product="Pristinamycin I synthase 2"
/protein_id="CAA72311.1"
/db_xref="GI:2052248"
/db_xref="GOA:O54959"
/db_xref="SPTREMBL:O54959"
/translation="MLVFDARLPLTAQGLTWFAQARDPHSPVENTGEVVELTGLD
ADRFPAALRATVARTDALRLRPHRTDAEGTIVTQSVADDVFTLHRVDVSDEDDPAAAD
AMRDVIARPLDVTCTGRLFTQALFTLGFGRHWFQHVHLLLDGFGFLVAERVAAD
VRAGAEAGDAEAGDTEPGIRDLIAEADWRSDAYAQDRAHWLATPRHTPTVLGTTA
PCAGALRRTASAGPEVLADAVARAATGQVAFPAAVAVVWLTCGERTVTLGMP
VTARTISAARRAPGTVANVPALHVRPRPAAPLAEULVASVRDALEPLLRHORYPEBAR
RELGAVERAYGTVNNWMPYGRALDFGFPAAASVHPVSTGTVEDLEKSVSEPDGSGAR
IDLEANAAYGESETDAHLARVFLGFAEAGPDPTLGRHAAPAAAPAAAPAAAPTGG
PALHTLFAARAATPHAVAGDGHDLTYAQLDARANLAHLRVEOGAERLVALTL
PSLDTLVALLAVALKSGAAYLPDDPYPADRLAMLEDAAPVLVLDPOQAMTDLGYP
DTAPRTAVDGAHPAVIYTSGSTGRPKGVVIPSNNVRLFTSTDHWFSGPDVWTLF
HSVADFSTWEIHWGALLHCGRLVVPVHVSRSFGPDLDLAREKVTVLNQTPTAFHQL
DAADRARTAAPEALARYVFGGEALDVAELADWAPRGTAARLVNMGITETTVHVT
ALPGHGVPTGSLGGFIPLDTPRVLDAALRPVPPGFTGELYVYAGAGLARGYLRP
ALTAPFPADPYGAPGRMYRTGDVLVRHLDGTYAYLGRGDQDVQKIRGRFRELGEIEN
VIATHEGVAQAANVGRDRHGLRLAAAYVPTGTPDVGWRALLAGQLPAHSLPAS
FTLLAALPLTANGKLDRAALPAQAPAPETADPADPGEAALRRIIADTLGTGRV
GADSDFTLGGHSLLAGLIIAAVRDTLGAEITIQDLFEAPTAGLAARVTRATGRPP
LYPRHGDGPRPTSVAQHWMVLRVEGAAPTNNMELARLSCTLNRTLRAALYDVV
VRHESLRTYPRGEHGVTRLLLPPEQARWVIESRVDDQLPAALTAAREPENLDCQ
LPIRAHLFTPGGDSVLLIMHHVAGDWSLRPLLDGLSLTAARLAARAPDPPLPV
SVADFVWQRELIGDEHDSLAARQALACHWLQAAALLTRIGAGTDTPICGPTASRPDS
SVPLAVPPQHLNRDFTFAHAGTACHWLQAAALLTRIGAGTDTPICGPTASRPDS
ADLDLVGFVFWPLVRADTSGDPTFTELLARATALAAHQDVFTRVVERLQPER
SAAHPLFLQVLLSYQDFEARHLKGLDVRREDVDFGVAKFDLTFLNLVHRGDGEFAG
```


QY 1152 LeuProLeuPro-----MetAspProLysLys 1160
Db 17040 GACCGCTGCCGACGACCTGGGCGCTACCGGAGACCGACCTACCGACCGCGACCGC 17099
QY 1161 LeuAlaProPheSerGlyValLysGlnGln-----LeuSerProArgGly 1176
Db 17100 ACGGCGCG-----CTACGGCCGACACCGCGGTACGTATCCACACCTCCGGC 17150
QY 1177 GlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGlnAlaSerValLeuArg 1196
Db 17151 TCCACCGGCACCCCAAGCGCGTCTCATGCCCGACCGCGCGCTGGCCAACTGCTGACC 17210
QY 1197 GlyThrAlaLeuGlySer-----ValProGlyGlySerIleThrLysGlyLeuProse 1214
Db 17211 TGGCACCGCGCGCTTCCCGCGCGCACCGGGGTGCGCA-----CCGCC 17255
QY 1214 rThrArgValProSerAspSerAlaIleThrTyArgGlySerIleThrHisGlyThrPr 1234
Db 17256 CAGTTACCGCCATCGGCTTCGACTTCTCGGTGACGAGATCTCTCCCGCTCGTATG 17315
QY 1234 oAlaAspValLeuTyLysGlyThrIleThrArgIleGlyGluAspSerProSerAr 1254
Db 17316 GGCA-----AGACCTTCGCC 17330
QY 1254 gLeuAspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyArgGlyLysLys 1274
Db 17331 GTGCCCTCGAAGAGGTCCGCCACAGCGCGAAC-----TGTCGGCGCGTGGCTC 17381
QY 1274 sGlyHisValLeuSerTyArgGlyGlyMetSerValThrGlnCysSerLys-----G1 1292
Db 17382 GAGACACGACATCA-----ACGAATGTCGCGCGAACCTC 17420
QY 1292 uAspGlyArgSerSerGlyProProHisGluThrAlaAlaProLysArgThr--Tyr 1311
Db 17421 GTCATCGAGCCCTCGCGGAGCGCGCGAGCGCGCGCCGCGCATCTCGCGGACCTCAC 17480
QY 1312 AspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMet 1331
Db 17481 GACATCTCCAGGG-----GCGAGGCGCTCGCCCCACCGAAGGTTACGCGCTTC 17534
QY 1332 GlyArgAlaIleProProGluArg---HisSerProHisLysLysGluGlnHisHis 1350
Db 17535 ACGGCGCGCTACCGCGCGCGCTCTGCAACGTGTACGGCGCGCGAGACCGACGCC 17594
QY 1351 lLeArgGlySerIleThrGlnGlyIleProArgSerTyValGluAlaGlnGluAspTy 1370
Db 17595 GTCACCAACCCACACCTCGCGCGCTGATCCCGCCACTGG-----17633
QY 1371 LeuArgArgGluAlaLysLeuLeuLysArgGlyThrProProProProProProSer 1390
Db 17634 -----CCCCCTCCGACCATCGGC 17654
QY 1391 Arg-----AspLeuThrGluAlaTyLysThrGlnAlaLeuGlyProLeuLysLys 1408
Db 17655 CGGCCCTGCACGACGACCGGTGTAGTCTTGAGACAGC-----GCATGCGC 17702
QY 1409 ProAlaHisGluGlyLeuAlaAlaThrValLysGluAlaGlyArgSerIle-----1425
Db 17703 CCGCTCCCGCGCGGTACCGCGAGCTGTACTGCGCGCGCGCGCGCTCGCGCGCGC 17762
QY 1426 -----HisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaPro 1443
Db 17763 TACTGACCGC-----CCGCGACACACCGCGAGCGTTCGTGCGGACCCCTACGCCCCA 17819
QY 1444 ArgPro-----LeuLysGluGlySerIle-ThrGlnGlyThrProLysLys--1458
Db 17820 CGGCCCGCGCGCATGTACCGACCGCGGACCTCGCGCGCTGGAAACACCGCGCGAG 17879
QY 1459 ----TyAspThrGlyAlaSerThrGlySerLysHisAspValArgSerLeuIl 1477
Db 17880 CTGGAGTTCCGCGCGCGCGACCCACAGGTCAAGATCC-----CGCGCTTCGC 17930

QY 1477 eGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaAr 1497
Db 17931 GTGGAGCGCGGAGGTTCAGGCGCGCTTCACCGCCACCGCGATCACCAGGAGCG 17990
QY 1497 gAlaLeuGluArgAlaCysTyArgGluSerLeuLysSerArgProGlyThrAlaSerSe 1517
Db 17991 GTCC-----TCGGCAGCGGACCGCTCGTC 18017
QY 1517 rSerGlyGlySerIleAlaArgGlyAla-----ProVa 1528
Db 18018 GCCTAGCTGTTCACCAACCGCGGACCGGACCTGACCGGGGTACCGGAGCACCTGGCGGCC 18077
QY 1528 lIleValProGluLeuGlyLysPro-----ArgGlnSerProLeuTh 1542
Db 18078 CGCTCCCGAGCTTATGTGTCCCGCCCGCTAGCTCGGCTCGAGGCGGTG--CGGTCTAC 18136
QY 1542 rTyArgAspHis-----GlyAlaProPheAlaGlyHi 1553
Db 18137 CGTGAACGGCAAACTCGACCGGCGCGCTTCCCGCACCGGACCGCGCGCGCGCG 18196
QY 1553 sLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySe 1573
Db 18197 CGCCCGCGC-----ACGTGCGGAGGA--GATCTGTGCGGCTTTCGCCG 18243
QY 1573 rLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAl 1593
Db 18244 AGTGTCTGGGACCGGCGCGCATCGACAACTTCTTCGACCTTCGGCGGCACT 18303
QY 1593 aLysSerProHisSerThrValProGluHisHisProHisProIleSerPro-----Ty 1611
Db 18304 CGTGTCTGCCAC-----CCGCTCATCTCCCGCATCCGACCGCTTCGGCGCGC 18354
QY 1611 rGluHisLeuLeuArgGlyValSerGlyValAspLeuTyArgSer-----1626
Db 18355 AGATCACTCCGCGGACCTGTTTCGAGGCGCCACCGCTCCGCGACTCGGGAGCGCTCG 18414
QY 1627 -----HisIleProLeu--AlaPheAspProThrSerIleProArgGlyLeuProL 1643
Db 18415 ACAGCGGTCCGGCGCGCGCTCCTCACCCTCCCGGAGCGCCCGGAGCGCTCCCGC 18474
QY 1643 euAspAlaAla-----AlaAlaTyTyLeuProArgHisLeuAlaProAsnProThrT 1661
Db 18475 TGTCCTACCGCAGCGCGCTGTGTCTTCGGCGCTCCTCCGCGCGGACCGCGCT 18534
QY 1661 yzProHisLeuTyProProTyTyLeuIleArgGlyTyTyProAspThrAlaAlaLeuGluA 1681
Db 18535 AC---AACATGCCGTCGAGCTCGGCTTCGCGGACCTTCGCGGACCGCGCGCTG---18587
QY 1681 snArgGlnThrIleIleAsnAspTyTyIleThrSerGlnGlnMetHisHisAsnThrAlaT 1701
Db 18588 -----ACCGCGCGCTTCGCGGACCTCTGTGGCC-----CGCCAGAAACCTGTC 18630
QY 1701 hrAlaMetAlaGlnArgAlaAsp-----1708
Db 18631 GCACCGCTTTCGCGAAGCGAGCAGCGGAGCGCGCCACCGAGCTCTGCTCAGCCCGCGC 18690
QY 1709 --MetLeuArgGlyLeuSerProArgGluSerSer-----LeuAlaLeuAsnTyRA 1725
Db 18691 CGCCCGCGCGCGCTGCAACCGCGGAGCACACGAGGAGCACCTTCGCCACGGAACCTGG 18750
QY 1725 laAlaGlyProArgGlyIleIleAspLeu---SerGlnValPro-----1738
Db 18751 CGCCCGCGCGCGCTTCGACTTCGACTCACACCGAATGCGCTGCGCGGCACTCTGT 18810
QY 1739 -----HisLeu--ProValLeu 1743
Db 18811 TCGCGCTCGCGAGGACGAGACGCTCTGTCTCTGTGTGTGACCAACATCGCGCGGAGC 18870
QY 1744 ValProProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyLeuProThr 1763
Db 18871 GCTGTGCTCGCGCGCGCTTCGCGCGGACCTGTGTACCGGCTACCGGCTTCGCGCGCGC 18930
QY 1764 AlaProGlnProPheSerSerArg---HisSerSerSerProLeuSerProGlyGly---1781

Db 18931 GCCACGCCCGCACTTCGCGCGCTGCGCGTCCAGTACGCGCACTACACCTGTGGCAGC 18990
Qy 1782 ---ProThrHisLeuThrLysProThrThrThrSerSerGluArgGluArgAspArg 1800
Db 18991 ACGCCCTCTCCGCGCAGACAGACACCCGACACCATCATCAGCGCAGAGCCCACT 19050
Qy 1801 AspArgGluArgAspArgAspArgGluArgGluLysSerLeuLeuThrThrThrThr 1820
Db 19051 GGAAGGACGCCCTCGCGCGCTGCGCGGAGGATCCAGC----- 19089
Qy 1821 ValGluHisAlaProLleTrpArgProGly----- 1830
Db 19090 -----TCCCGCGCAGCGCGCGCGCGCGCCGCCACCCCGCACACCGCGGGAACCG 19140
Qy 1831 -----ThrGluGlnSerSerGlySerSerGlySerSer 1841
Db 19141 TCTTCGCCACCTCGACGCCGACCTGCACCGCAGCTGTGCACCTCGCCGCGACCGCG 19200
Qy 1842 GlyGlyGlyGlySerSerSerArgProAlaSerHisSerHisAlaHisGlnHis--- 1860
Db 19201 GCGCCAGCCTTTCATGCTCTCCAGCGCGGTGCGCGGTACTCACC-CGSCACGCG 19259
Qy 1861 -----SerProIleSerProArgThrGlnAspAlaLeuGln 1872
Db 19260 GCGGACCGACATCCCGCTCGGACGCCCGCATCCCGCGCGCACCGACGCGCTCGAC 19319
Qy 1873 GlnArgProSerValLeuHisAsnThrGlyMetLysGlyLleIleThrAlaValGluPro 1892
Db 19320 GACCTGTGGCTTCTTCGTCAACACCTCTGCTGCGCAGCGACCTCCGGCAACCCC 19379
Qy 1893 SerLysProThrValLeu---Arg-SerThrThrSerSerProValArgProAlaAl 1911
Db 19380 ACCTTCGCGAATCTCTCGACCGGTACGCGAGTTCGACCTCGCGCTACGCCACAC 19439
Qy 1911 aThrPhePro-----ProAlaThrHisCysProLeuGln 1922
Db 19440 GACGTGCTCTCGAAGCTCTGAGGACGCGCGCACCGCGCGCTCTCTGCGCGCAC 19499
Qy 1922 YGlyThrLeuaspGlyValTyProThrLeuMetGluProValLeuLeuProLysGluAl 1942
Db 19500 CCGCTCTTCAGGTCTGCTGCGCCCTCGAGAACACCCCG-----AA 19541
Qy 1942 aProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLy 1962
Db 19542 GCGCTCTCGAATCGCGGACTCACCGCGCGCGCGCGCGGTCTGCTACCC-----TCGGC 19595
Qy 1962 sProProAla-----ArgSerGlyLeuGluPro----- 1971
Db 19596 GCGGCCAAGTTCGACCTCACCTTCACTGCGCGAAGCGCACCGCGCGCGCAACCC 19655
Qy 1972 -----AlaSerSerProSerLysGlySerGluProArgProLeuValProValSerGln 1990
Db 19656 CACGGCATCAGCGCGCTCGAATACGACACCGACCTGTTCGACCGCACCGCCGAA 19715
Qy 1990 yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAs 2010
Db 19716 GCACTCACCGACCGGCTCATCACT-----TCTGCGCGCG-----TCGCGCGC 19760
Qy 2010 pProProAlaProProAlaSerAla-SerAspProHisArgGluLysThrGlnSerLysP 2030
Db 19761 GACCGCGCGCGCACCGTTCGCGAGTTCGCGATCTCTCGACGCGCGACCGACCGCGGTC 19820
Qy 2030 roPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyHisGlySerSerTySerP 2050
Db 19821 CTACCGGCTGGAACGA-----CACCGACACCCCGCGCGC 19856
Qy 2050 roGluGlyValGlu-----ProValSerProValSerSerProSerLeuThrHisAspL 2068
Db 19857 CGGCGCACCGCGCGCTGCTCGCGCGCGCATCCCGACAGCGCGCGCGCACCGCG 19916
Qy 2068 ysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeuGluGlyLeuArgp 2088
Db 19917 CACGCGTCCGCTACCCGAAGCGCGCGCGCACCTGTATACCTACACGAACTCGACGCC 19976
Qy 2088 roLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgp 2108
Db 19977 CGGCGCAACCG-----C 19988
Qy 2108 roLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaPro-GlyValLys 2127
Db 19989 CTGCGCGCGCACCTGTATCCG-----CCGCGCGCTCACC 20021
Qy 2128 GlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTy 2147
Db 20022 GCGGAGACCGCGTTCGCGCAACGCTCGCGCGCTCG----- 20068
Qy 2148 ThrArgHisHisProGln-----GlnLeuSerAlaProLeuProAlaProLeuTySer 2165
Db 20069 ---CGTCAACCACTCGGATCTCAAGCGCGCGCGGTGTACGTGCGCTGCACACCGG 20125
Qy 2166 Phe-----ProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp 2182
Db 20126 CTATCCGCTCGACCGGATCGGACGCTCTCGCGCACACCGAGCGCGCTGTCTGTCTAC 20185
Qy 2183 LeuTyrlLeuPro-----ProProAspHisGlyAlaProAlaArgGlySer 2197
Db 20186 CGACACCCACCGCGCGCCACCGCGCGCTCGGACACCCCGCGCTCACCGTCGACGA 20245
Qy 2198 ProHisSerGluGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGly 2217
Db 20246 GGACACCA-CGGCGCGCGACGACACACCGCGC---CGGAGTCACTCCGTCGCGC--- 20298
Qy 2218 GlyGluaspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSer 2237
Db 20299 -----CGGACCACTCGCTCATCATCTGTTACCTCCGCGCT--- 20334
Qy 2238 ArgSerAlaValTyProLeuLeuTyArgAspGlyGluGlnThrGluProSerArgMet 2257
Db 20335 -----CGACCGCGCACCCCAAG--- 20352
Qy 2258 GlySerLysSerProGlyAsnThrSerGln-----ProProAlaPhePheSerLysLeu 2275
Db 20353 GCATCGCATCACCCACCGCGCGCCATCGCGCTGCGCGCGCGCGCTGTGGACCTGG 20412
Qy 2276 ThrGluSerAsnSerAlaMetValLysSerLysGlnGluLeuAsnLysLysLeuAsn 2295
Db 20413 ACACCGCTCCCGCTCTGATGCACTCCCGTACGCTTCGACATCTCCACCTTCGAAC 20472
Qy 2296 ThrHisAsnArgAsnGluProGluTyArgAsnLysSerGlnProGlyThrGluIlePheAsn 2315
Db 20473 TGTTCGCGCGCTGCTGCGCGCGCGGATCTGCTG-----TCG 20511
Qy 2316 MetProAlaIleThrGlyThrGlyLeuMetThrTyArgSerGlnAlaValGlnGluHis 2335
Db 20512 CACCGCGCGGCAATCGACGCG-CGCGCTCGACGCGCACCTCGCGCGC-----CAC 20564
Qy 2336 AlaSerThrAsnMetGlyLeuGluAlaIleArgLysAlaLeuMetGlyLysTy--- 2354
Db 20565 GCGCTCACCTCGCTGCTGCTGACCGC-----GGCTGCTCGCGCTGATCGCC 20612
Qy 2355 AspGlnTrpGluGlu-SerProProLeuSer-----AlaAsnAl 2367
Db 20613 GACGAGCACCGGAGGTCTTCACCGCGCTCAAGACGCTGGACCGCGCGCGCTGCTC 20672
Qy 2367 aPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
Db 20673 TCTCCCGACCGCTGCGCGCTCTCGAAGCTGCGCGCGCACCG----- 20718
Qy 2387 pGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGln 2407
Db 20719 -----TCGTC 20723
Qy 2407 yArgProSerSerArgLysAlaLysSerPro-----AlaProGlyLeuAlaSer----- 2423
Db 20724 AAGACCTCTACGCGCGCGACCGACACCCCTCGGCTGCGCTGCTGCTGCTTACCGCAC 20783

GERLDQASGARPALTPAERPERLPLSHAQRRLWFLGRLDGNSTYNIPLALRLSGTLD
TAALRALADLVARHETRLTRYPSEHDEQYCHIVAPADATPALTVVTHREHLEAQLA
EATAPYDVLTEPRLTRFRIEGLDEHVLVLLHLLHAGDWSLAPVTRDLAAAYAAAR
DGHAPGAPLPVOYVYTLWORGILGDADDPSTLOAAQLAHWRQALEGRPAHLELPCD
HBPVATHRGATVPHEHIDAGLEKLTALAKACDSSLFWLQAAPALTRHGAGDPI
PVGSPVAGRTDADLDDLVGFVNTVLVRLTUTSGDTPPRELVARVQFDLAAYTHQDMP
FEKLEBVENPERSLARNPFLQVVALQSPMAADLAUPLGLDVAAPVRVPGAKFDGLIA
VIEBHATDGTFRAGIRDMWSTELFERGTVEALGSLRLLDADVPDPTIGSVOLL
DATERHLLVGLANTDPLPEERTLTALFEQAAATPDATALVWGRSRLTYAELDARA
NRLARHLVAGGPEQIVALQPLRSILDVTLAVLAWKSGAAYLPVDPDYPPIAHMR
ADAPALVDAI PDDTTLAAYADSRLLTDASAPLPAHPAYVITYSGSGTAPKGVV
AHSLSATVPAQAAGFGTHSRVINFASISFDAVWELTSLAGTGLVADADULL
PGPSLARLHRRHITLALPSPALPDPGALPGTDLIVAGDATAPDQAARFAPGR
MVNAYGLTETVTCATSEPATGDCAPP IGRPVAHARVYVLDRLRPVPGVTGEMTYS
GAGVAGYHTRPALTAQRPVDPYALLFGSTGTMYRTGDLARLDADGRHLPAGRADQ
QVKIRPRIEPGEIETVLTAAHPAAGAVIAREDTGDKOLVAVLTDTHHAAPDQD
LGTCQVDRWQOTDLSLYAEPGAFGEDFSGNSSYTAERIPLIEMREWRPAATVDRVL
ALAPRVLETIGCGTGLISQVAPHESTEIRGTDLSRSVARLAHLARGPLADKVTVD
ARAHEHDDLPKGRFDITVLNSVAQYPPDARYLAGILHRAAELLAPGTFILGDIRNL
RTLHAFTVLELRAGFADPAARVRAVEQSALATEKELLDDPFDLVDRLPHLTRA
DITLRTAAHNENSRHRYDVLTHRGDSAPAEVTLRWGTDLDTAGELARLVLSGRPV
RVGTINARLITGETRAQAALERSAADACAPALAPAGALDPHDVVALGAVCTWGD
DDTADAYAGGTAYRPTGQGPLANDPARAEQDAAALVGBELDLVQRLPAHMAPAY
VLDRPLSANGKLDALPAPDRGDAQAPTRREELSTLFAEVLGIPAVGDIR
SFDLGHSLLLATLSRISIRSVLGAELAVRDLFOAPTAAALAEVVGAGQAPPLTAR
TRAEVPLSPAQYRLWFLHMEGPGFTYNI PMSLRITGLTDLTAAALAAADVTARHEA
LRTVYPSGQVPRQHVLDHVRPHFTADTTAQTLEADLTAAARRGPDALRELPLRATL
FRLAEDEHVLILLMHITAGDQSWPPLARDLDGAYAAARLAGRAPDLPLPVQVADVAL
QDRLDGEDDDPTSRHARQLAYWAGALDDLPESELQPLCDRPRPAVATHRGDTVPFVAD
PDHQRNLATSGRASLFWLQAFAALLTRHGAGDIPLGSPFAGRTDUALDEDLVG
FFVTLVLTDTGSDPTFRELQVRRESIDLAAYAHQDVPEFELVEXLRPERSLSHPL
FQVMLAFQNSGEARLPLDGLARALPVGVAKPFDLHLSNVELRTDGRPGGLOAALE
YATDLDFRTAQALVRLRLAAVADDPGLTVGRVELLDPAERHRLPDWDMTPGCD
GAARTPVTELEQOARTPDVALVHDDGRLTYAEHLHARANRLHLLTNGAPGEQIV
ALMRPSLDLYVALLAVLTKAAYLPVDISYPAERIAPMEDARPVTLDRLPDDLGA
VRDUTDADRTAPRPEHPAYVIHTSGSTGTPKAVPMHAGLANLLTHWARFPQGT
GVRTAFTAGTFQFQVEILSPVMGKTLAVPSEVRHSAELLAGLWETQOINELFAP
NLVIEALAAEABAGRTLPLDITLOQGEALAPTRVAPTAAPVGRHNVYGPART
HAVTHTLAADPAHWPSPAPIGRPVDRVVLDSALRPVPGVTGELYLAGAVARG
YLHRPRTAERFVADPVAPRGARMTVTDGLRWNTGEGLEFAGRADHOKVIRGFVE
PGEVEALTPRITQAAVLHAGDRLVAVYVTTADRDLTGVRHLEALRLPDMVPTAY
VRUEALPTVNGKLDRAAPDAPAGRAPRTVREILGCLFAEVLGTERAGIDNFP
FDLGHSLALTRISIRLTALGABITLRLDLEAPTAVAGLERLDQASGARPALTPAER
PERLPLSQAORRLWFLRGDGYNMPVGLRSLGTLDTAALTAALADVAARHETLR
TVPEGDSGPHOLVLPDGAARPALPRDTPDQDLATLAAAHCHFDLTTELPLRAT
LFLRAEDHVLILLHIIAGDWSLAPLARDIATAAARDGHAPDFAPLPVQYADYT
LWQHALLGDENDPSTISROSAHWKALAGLPEEIQLPADRPRPAHARGEAVFTIL
DADHRELLHARTGASLFWLQAFAVLTIRHGAGDIPLGSPFAGRTDADLDDLV
GFFVNTLVLTDTSGNPTFRELQVRRESIDLAAYAHQDVPEFELVDEDAAPGRSLRHP
LFQVLLALQNTBALVLELPGLTARPEVTLTGAARFDLTFLNLAERHGPDPGPHGIDAVL
EYSTDLDFRTTBAALTDRLITFLRAVAADPGRTVADVPILDADHARVLTGWNDDHP
RGRTADACLPRIIAQOARTPHAVATEAGGTLITYSELDARANRLHRLIRGWTYA
ETRVAVLAERSAQVLVTLTALLKAGVYVPLHTGYVDRMRHVLTADTEAALLTDTH
AATAARDTPALTVDTEOTAGEQDTPADTVRPPQLAYIMFTSGSTGTPKGIITHR
DATALAADRCMDLDSRVLHSPYAPDITSTFELFAPLAGRIVAVPRGDIDAVAL
RTLAAGHVTSLMTAGLLGVIADAEPEVFTGKDVMTGDSVSPVAVRVLACPGTV
VKTLYGTEFTLCTLPFTDPRRIIPAVPIGRPLDNTRAYVLDRLRPVPGVTGEL
YIAGALGRLVQDSQARTAEFTADPHALFGDTGGEMVFTGLARDDADGLVFCGR
ADQVKIRGRIRIEGEIETALAHPDVTRAARVAPRGADKVLVLYLVTAPDAGDTT
AEQURAGLETQLEPDMVPAFVALPALPFPNGKLDRLDALPEPDWGGAGRPGRPRE
EILCLGFAPDAPRYPTDNDFFELGHSMLATLVRKTVILGADIGVTLFEAPTV
AALAAIRDGADTHDPFGVVLPLRTGSRAPLFCVHPAGFGWVYSGALLRHTDREOPL
YALQAARGLARPELPDDIDAMARDYABQIRKTVPEGPYEILGWSFGSLVAHAAVRL
AEGAEVSLAVLOGYPDAYDGTETHEVEGEQVLAILLNAGVDRQAQFGDAPLQRAAVL
EKURDSGLNGLNDDHDAVGRVTVFLNTRKLIQNFRPRTGDMVTFPAAAGHHDHAL
TPGNWRPYVTGRIEHEHLLTDHAGLARPEALGAVARTLAQRDTPQAGH"

ORIGIN

Alignment Scores:

Pred. No.:	5,24e-05	Length:	22449
Score:	654.00	Matches:	683
Percent Similarity:	31.26%	Conservative:	290
Best Local Similarity:	21.94%	Mismatches:	1193

QY 254 LeuPro-----LeuTyrAsnGlnProSerAspThrArgGln 265
|||
D 13604 CTTCCGATCGAACCCGGGAGATCGAGACCGTCTGACCGCCACCAGCCCGCGTCCGCG 13663
QY 266 TyrHisGlnAsnIleLysLeuAsnGlnAlaMetArg-LysLysLeuIleLeuLeuPheLys 285
|||
D 13664 CGCGCGGTGATCGCGCGGAGACACCCCGGCGACAAACAGCTCGTCTACCTGAC 13723
QY 285 sArg-----ArgAsnHisAlaArg-----LysGln 293
|||
D 13724 GCGGACACCCACCCAGCCCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 13783
QY 293 nTrpLysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGlnLysVa 313
|||
D 13784 CTGGCAG-----CAGACCTACGACTCCCTCTACGAGCGGAAACCCCGCGCGC 13831
QY 313 l-----GluArgIleGluAsn---AsnProArgArgAlaLysGlnSerLysValAr 330
|||
D 13832 CTTCCGCGAGACTTCTCCGCTGGAACACAGCTACACCGCGGAGGATCCCTGGA 13891
QY 330 gGluTyrTyrGlnLysGlnPheProGluIleArgLysGlnArgGlnGlnGluArgMe 350
|||
D 13892 GGAGATCGGAGTGGCGCGCCGACCGTCCGACCGCGTCTCGCCCTGGCGCCCGCG 13951
QY 350 tGlnSerArgValGly----- 355
|||
D 13952 GGTCTGGAGATCGGCTGGCGACCGGCTGATCTCTCCAGGTGCGCCCGCCACACCGA 14011
QY 356 ---GlnArgGlySerGlyLeuSerMetSerAla----- 365
|||
D 14012 GGAGTACGGGGACCGACCTCTCCGACGGTCTGCGCAGGCTGGCGGCCACCTCGC 14071
QY 366 -----AlaArgSerGluHisGluValSe 373
|||
D 14072 CGGCGCGCCGACCTGGCGGACAAAGTACCGTGGACCGCGCGCGCGACACGAGCGGA 14131
QY 373 rGluIleLeuAspGly-----LeuSerGluGlnGluAsnLeuGluLys 387
|||
D 14132 CGACTGCGCCAGGCGCTTCGACGATCGTCTCAACTCGTCCGCGCGCTACTTCCC 14191
QY 387 sGlnMetArgGlnLeuAla-----ValIleProPro----- 397
|||
D 14192 CGACGCGCGTACTCGCGGATCTCTCCACCGGCGCGCGAACTGTCGCGCGCGCGG 14251
QY 398 -MetLeuTyrAspAlaAspGlnArgIleLysPheIleasn----- 411
|||
D 14252 CACGATCTCTCGCGGATCTCCGACCTCGGACCGCTCGACCGCTTCGCGCGCGCGT 14311
QY 412 -----MetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnVa 428
|||
D 14312 CGAACTGCGCGCGCGCGCTTCGCGGACCGCGCGCGGCGTA-----CGCGCGC 14362
QY 428 lMetAsnMetTrpSerGluGlnGluLysGluThr-----PheArgGln 442
|||
D 14363 CGTCAACAGTCACTGGCGACGAGAGGAATCTCTGTCGACCGCGGACTTCTTCAACGA 14422
QY 442 uLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysTh 462
|||
D 14423 CCGTGTGACCGCGGACCGCGAC-----CTGACCGCGCGCGGACATCACTCCGCGCAC 14473
QY 462 rValAlaGluCysValLeuTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLe 482
|||
D 14474 GCGCGCCAC-----CACAAACGAGATGAGCGCGCACCG 14506
QY 482 uValArg----- 485
|||
D 14507 -CTACGAGCTACCTCCACCGCGCGCGGACAGCGCGCGCGCGGAGACCGTCTGCG 14565
QY 486 -----SerTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGln 500
|||
D 14566 GCTGGGCGACCGACTTCACACCGCGCGGAACTCGCGCGGTCTCTCCGCGCGCGC 14625
QY 500 nGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerSerGlnGluGlu 520

D 14626 CCGTGGCGCTCACCGGATCCCAACGCGCGCTCACCGGAGACACGCGCGCACAAAG 14685
QY 520 sAsp-GluLysGluLysGluLysGlu-----AlaGluLysGluGluGluLysProGluV 538
|||
D 14686 CCGTCGAACCGGATCGCGCGCGGACGCTCGCGCGGCTCGAAGCCCCCGCGCGCGC 14745
QY 538 alGluAsnAspLysGluAspLeuLysGluLysThrAspAspThrSerGlyGluAspA 558
|||
D 14746 GCGCACTCGACCCCGACGAGCTGTACGCGCTGGCGCGCTGTGCACCTGGGCGCACGAG 14805
QY 558 snAspGluLysGluAlaValAlaSerLysGlyArg-LysThrAlaAsnSer---GlnGly 576
|||
D 14806 ACACACCTTCGACCGCTACGCGCGGACGCGCTACGCGCGCGCGCGCGCGCGC 14865
QY 577 ArgArgLysGlyArgIleThrArg-----SerMetAlaAsn 588
|||
D 14866 AGGAGCCCTCGCCACACCGCGCGCGCGCGGCGGACGAGCGCGCGCTGTGCGCGAAC 14925
QY 589 GluAlaAsnSerGluGluAlaIleThrProGlnGlnSerAlaGluLeuAla----- 605
|||
D 14926 TCGCGGACTCTGTCCCGGAGCG-CTGCCCGCCACATGGCGCGCGCTACGTCCTG 14984
QY 606 -----SerMetGluLeuAsnGluSerSerArgTyrThrGluGluMetGluThrAla 623
|||
D 14985 CTCGACCGCTCGCGTACGCGCAACGCGCAACTCGACCGCGACGCGCTGCCCGCCGCC 15044
QY 624 LysLysGlyLeuLeuGluHisGlyArgAsn-----Trp 634
|||
D 15045 GACCGCGGAGAGCG 15104
QY 635 SerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPheTyrPhe 654
|||
D 15105 ACCCTGTTCCGCGAGTGTGCGCTGCGCGCGCTCGCATCGACCGCGCTCTTCGA- 15163
QY 655 AsnTyrLysLysArgGlnAsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGlu 674
|||
D 15164 CTTGGCGCGCACTCGTTCGCGACCGCGCTGTCTCAGCCGCTCGCGCGCTCTCTGGG 15223
QY 675 LysGlu-----ArgAsnAlaArgArgLys 682
|||
D 15224 CGCGGACTCGCGCTCGCGACCTGTTCCAGCGCGCGCGCGCGCTCGCGCGCGCGCG 15283
QY 683 LysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValGluAsp 702
|||
D 15284 CGTCGACG 15310
QY 703 GluGluMetGluAlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGlu 722
|||
D 15310 ----- 15310
QY 723 AlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAlaThrVal 742
|||
D 15311 GCGCTCAACG-CCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT- 15351
QY 743 AsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLysAspThr 762
|||
D 15352 -----CCTTCG 15384
QY 763 GlyGluAsnGlyProLysProAlaThrLeuGlyAlaAspGlyProProGlyPro 782
|||
D 15385 GCATGAGGCG 15444
QY 783 ProThrProProArgThrSerArgAlaProIleGluPro----- 796
|||
D 15445 TGGACACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15504
QY 797 -----ThrProAlaSerGluAla-----ThrGlyAla 805
|||
D 15505 GCACCGTCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15564
QY 806 ProThr-----ProProAlaProProSerProSerAlaPro-----Pro 819
|||

Qy	1371	LeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProProProProSer	1390
Db	17634	-----	17654
Qy	1391	Arg-----AspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLys	1408
Db	17655	CGCCCGTGCACACACGCGGTGTACGTCCTGGACAGC-----GCACTGGCG	17702
Qy	1409	ProAlaHisGluGlyLeuValAlaThrValLysGluAlaGlyArgSerIle	1425
Db	17703	CCCGTCCCGCGCGGTACCGCGAGCTGTACTTGGCGCGCGCGGTCCCGCGCG	17762
Qy	1426	-----HisGluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaPro	1443
Db	17763	TACCTGACCGC---CGCGCACACCGCGAGGTTTCGTGCGCGACCTACGCCCA	17819
Qy	1444	ArgPro-----LeuLysGluGlySerIle-ThrGlnGlyThrProLeuLys---	1458
Db	17820	CGCCCGCGCGCGCATGTACCGCACCGCGAGCTCGCGCGTGTGGAAACACCGCGCGAG	17879
Qy	1459	----TyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspValArgSerLeu1	1477
Db	17880	CTGAGTTCGCGCGCGCGCGCGCACACAGTCAAGATCC-----CGCGTTCGCG	17930
Qy	1477	eGlySerProGlyArgThrPheProValHisProLeuAspValMetAlaAspAlaAr	1497
Db	17931	GTGAGCGCGCGAGGTGAGCGCGCGCTCACCGCCACCCCGCATCACCCAGGCGG	17990
Qy	1497	gAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSe	1517
Db	17991	GTCC-----TGCGCGACGCGCGCGCTCGTC	18017
Qy	1517	rSerGlyGlySerIleAlaArgGlyAla-----ProVa	1528
Db	18018	GCCTACGTCTACACCGCGCGGACCGGACCTGACCGGGTACGCGAGCACTGCGCGC	18077
Qy	1528	lIleValProGluLeuGlyLysPro-----ArgGlnSerProLeuTh	1542
Db	18078	CGCTGCCGACATTCATGTTCCCAACCGCTACGTCCGCTCGAGGCGTG-CGGCTCAC	18136
Qy	1542	rTyrGluAspHis-----GlyAlaProPheAlaGlyHi	1553
Db	18137	CGTGACCGCAAACTCGACCGCGCGCGCTGCCCGCACCGGACCGCGCGCGCGCG	18196
Qy	1553	sLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlySe	1573
Db	18197	CGCCCGCGC-----ACCGTCCGCGAGGA-GATCCTGTCCGCGCTGTCGCGC	18243
Qy	1573	rLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAl	1593
Db	18244	AGTGCTGGGCACGCGCGCGGATCGACGACAACTTCTCGACCTCGCGGCCACT	18303
Qy	1593	AlaSerProHisSerThrValProGluHisHisProHisProIleSerPro-----Ty	1611
Db	18304	CGCTGCTCGCAC-----CGCGCTCATCTCCCGCATCCGACCGCGCTGGCGCGC	18354
Qy	1611	rGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSer-----	1626
Db	18355	AGATCACCTGCGCGACCTGTTGAGCGCGCCACCGTCCGCGGACTGGGGAGCGCTCG	18414
Qy	1627	-----HisIleProLeu-AlaPheAspProThrSerIleProArgGlyIleProL	1643
Db	18415	ACAGGCGTCCGGCGCGCGCTCACCCCGCGAGCGCCCGAGCGCTCCCGC	18474
Qy	1643	euAspAlaAla-----AlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrT	1661
Db	18475	TGCTCTACGCGCGCGCTGTGTGTTCTCGCGCGCTCACCGCGCGGACCCCGGCT	18534
Qy	1661	YrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluA	1681
Db	18535	AC---AACATGCCCGTCCGAGCTGCGCTGTCGCGCACCTCGACACCGCGCGCTG---	18587
Qy	1681	snArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlat	1701

Db 19596 GCGGCCAGTTCGACCTCCTCAACCTGGCGGAACGGCAGCGCCCGGAGAACCC 19655
 Qy 1972 ----AlaSerSerProSerLysGlySerGluProArgProLeuValProProValSerG1 1990
 Db 19656 CACGGCATCGACCGCGTCTCGAATACAGCAGCAGCAGCTTCGACCGCAGCCACCGCGGAA 19715
 Qy 1990 yHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAs 2010
 Db 19716 GCACCTCACCGACCGCGTCTACACT-----TCCTGCGCGCGC-----TCGCGCGC 19760
 Qy 2010 pProProAlaProProAlaSerAla-SerAspProHisArgGluLysThrGlnSerLysP 2030
 Db 19761 GACCCCGCGCGACCGTGGCGGAGTCCGATCTCGACCGCAGCAGCAGCGCGGTC 19820
 Qy 2030 roPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerP 2050
 Db 19821 CTCACCGCTGGAACGA-----CACCGACCAACCCCGCGCGC 19856
 Qy 2050 roGluGlyValGlu-----ProValSerProValSerSerProSerLeuThrHisAspL 2068
 Db 19857 CGCGCGCACCGCGCTGCTGCTCCCGCGCATCGCCGAACAGCGCGCGCGCACCGCCG 19916
 Qy 2068 yGlyLeuProLysHisLeuGluLeuLysSerHisLeuGluGlyGluLeuArgP 2088
 Db 19917 CACGGGTCCCGTCCACGAAGCGCGCGGACCCCTGATCATCTACAGCAACTCGACGCC 19976
 Qy 2088 roLysGlnProGlyProValLysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgP 2108
 Db 19977 CGCGCCACCG-----C-----C 19988
 Qy 2108 roLeuProGluSerGlnProSerSerSerProLeuLeuGlnThrAlaPro-GlyValLys 2127
 Db 19989 CTGGCGCGCACCTGATCCG-----CCGCGCGGTCAACC 20021
 Qy 2128 GlyHisGlnArgValThrLeuAlaGlnHisIleSerGluValIleThrGlnAspTyr 2147
 Db 20022 GCGGAGACCGCGTCCGCTCGCGGAACGCTCGCGCGAGTCG----- 20068
 Qy 2148 ThrArgHisHisProGln-----GlnLeuSerAlaProLeuProAlaProLeuTyrSer 2165
 Db 20069 ---CGTCACACCTCGGATCTCAAGCGCGCGGGGTGTAGTGCCTCGCACCGG 20125
 Qy 2166 Phe-----ProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp 2182
 Db 20126 CTATCCGTCGACCGGATCGGACGCTCTCGCGCACGACGAGCGCGCTGTGTCTCAC 20185
 Qy 2183 LeuTyrLeuPro-----ProProAspHisGlyAlaProAlaArgGlySer 2197
 Db 20186 CGACACCCACCGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20245
 Qy 2198 ProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGly 2217
 Db 20246 GGACACCA-CGGCGCGGACGACGACACCGCGC-----CGGAGTCAACGTCGCGG- 20298
 Qy 2218 GlyGluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSer 2237
 Db 20299 -----CGACACGCTCGCTTACATCATGTTCACCTCGCGCT----- 20334
 Qy 2238 ArgSerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMet 2257
 Db 20335 -----CGACCGGACCGCCCAAGG----- 20352
 Qy 2258 GlySerLysSerProGlyAsnThrSerGln-----ProProAlaPhePheSerLysLeu 2275
 Db 20353 GCATCGGATCACCCCGGACGCGCATCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 20412
 Qy 2276 ThrGluSerAsnSerAlaMetValLysSerLysGlnGluLeuAlaSerLysLysLeuAsn 2295
 Db 20413 ACACCGGCTCCCGCTCTGATGCACTCCCGTAGCGCTTCGACATCTCCACCTTCGAAC 20472
 Qy 2296 ThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsn 2315
 Db 20473 TGTTCCCGCGTCTGCG 20511

Qy 2316 MetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHis 2335
 Db 20512 CACCCCGCGCGCATCGACGCG-CGCGTCTCGCAGCGCACCTCGCGCGC-----CAC 20564
 Qy 2336 AlaSerThrAsnMetGlyLeuGluAlaIleAlaLeuArgLysAlaLeuMetGlyLysTyr--- 2354
 Db 20565 GCGGTCACTCTGCTGCTGCTGACCGCG-----GGCCTGCTCGCGGTGATCGCC 20612
 Qy 2355 AspGlnTrpGluGlu-SerProProLeuSer-----AlaAsnAl 2367
 Db 20613 GACGAGCAGCGAGGTCTTCACGCGGTCAAGACGTGGACCGCGCGCGCGCGTCTC 20672
 Qy 2367 apHeAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAs 2387
 Db 20673 TCTCCACCGCGTGGCGCGCTCTGAAGCGCTGCCCGCGCACCG----- 20718
 Qy 2387 pGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerG1 2407
 Db 20719 -----TCGTC 20723
 Qy 2407 YArgProSerSerArgLysAlaLysSerPro-----AlaProGlyLeuAlaSer----- 2423
 Db 20724 AAGACCTCTACGCGCGCGCGGACCGACCCCTGGGTGCGACCTGCTCGCGTTCACCGAC 20783
 Qy 2424 -----GlyAspArgProProSerValSerValHisSerGluGlyAspCysAsnAr 2441
 Db 20784 CCGCGCGCATCCCGCGCGCGTCC----- 20808
 Qy 2441 gArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPr 2461
 Db 20809 -----CCATCGCGCGCGCGCTGGACAACACCGCGCGCTACGCTCTCGACGACGCG 20861
 Qy 2461 oPheProTyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerPro----- 2479
 Db 20862 CGCGCGGTACCGCGCG-----CGGTACCGCGCGAG 20891
 Qy 2480 -----ProProProGlyLeuProAlaGlySerGlyProLeuAla---GlyProHi 2495
 Db 20892 CTGTACATCGCGCGCGCGCGCTCGCGCGGTACTGTGGACGACGAGCGCGCACCGCG 20951
 Qy 2495 sHisAla 2497
 Db 20952 GAACGCT 20958
 RESULT 100
 SC0939114
 LOCUS
 DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 11/29.
 ACCESSION AL939114 AL096852 AL138058 AL138851 AL138852 AL163641 AL163672
 AL353832 AL353862 AL355774 AL442165 AL499570 AL596030 AL596138
 AL645882
 AL939114.1 GI:24413773
 VERSION
 KEYWORDS
 SOURCE Streptomyces coelicolor A3(2)
 ORGANISM Streptomyces coelicolor A3(2)
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomycetes.
 REFERENCE 1
 AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,
 Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,
 Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,
 Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornby, T., Howarth, S.,
 Huang, C.H., Kiser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,
 Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S.,
 Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,
 Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrall, B.G.,
 Parkhill, J. and Hopwood, D.A.
 TITLE Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)
 JOURNAL Nature 417 (6885), 141-147 (2002)
 MEDLINE 2196410
 PUBMED 12000953

```

REFERENCE 2 (bases 1 to 313800)
AUTHORS Bentley,S.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hanxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT On or before Oct 26, 2002 this sequence version replaced
gi:5531349, gi:20520908, gi:6911971, gi:6912004, gi:20520853,
gi:12224900, gi:20520885, gi:7649483, gi:7799193, gi:20520703,
gi:14018350, gi:20520873, gi:14970931.
FEATURES             location/Qualifiers
   'source            1..313800
                       /organism="Streptomyces coelicolor A3(2)"
                       /mol_type="genomic DNA"
                       /strain="A3(2)"
                       /db_xref="taxon:100226"
   gene              182..1027
                       /gene="SCO2781"
                       /notes="SCO2781, hypothetical protein, len: 281 aa;
                       similar to TR:CA556689 (EMBL:AL121596) Streptomyces
                       coelicolor hypothetical 28.7 kD protein, 259 aa; fasta
                       scores: opt: 261 z-score: 302.6 E(): 2e-09; 34.9% identity
                       in 255 aa overlap"
                       /codon_start=1
                       /transl_table=11
                       /product="hypothetical protein"
                       /protein_id="CA887218.1"
                       /db_xref="GI:7544045"
                       /db_xref="SPTREMBL:Q9L073"
   CDS                182..1027
                       /translation="MRVPDPAPPAGKRRHPPSPALARVTFAGDPLAFRSDGDSLS
                       LFLPHQAPAPVPVRLGEGWGWRELPEADVAVMRVSLRLSDTQGHAEIVD
                       FVLHGEPDSGIOAGPAARWAADAAGDRVLLGCPAVANRAIRFRPPEDTLVWIG
                       DETAVPAACVAIVLPAFTGKARVQLQVPHAEVDQDLRTAADAEITVLVWGDAAGPEAT
                       LATLRAAQPPAEPHPYVWTAGSGCVKRLRRHFVGVGRVDRRTVTFVPRGLTTEQ
                       LREQG"
   RBS                1178..1183
   gene              1191..2633
                       /gene="SCO2782"
                       /notes="SCO2782, probable pyridoxal-dependent
                       decarboxylase, len: 480 aa; similar to SW:RHSB RHIME
                       (EMBL:AF110737) Rhizobium meliloti L-2, 4-diaminobutyrate
                       decarboxylase (EC 4.1.1.-) RhsB, 495 aa; fasta scores:
                       opt: 1111 z-score: 1326.2 E(): 0; 41.9% identity in 458 aa
                       overlap. Contains pfam match to entry PF00282
                       pyridoxal dec, Pyridoxal-dependent decarboxylase conserved
                       domain"
                       /codon_start=1
                       /transl_table=11
                       /product="putative pyridoxal-dependent decarboxylase"
                       /protein_id="CA887219.1"
                       /db_xref="GI:7544046"
                       /db_xref="GOA:Q9L072"
                       /db_xref="SPTREMBL:Q9L072"
                       /translation="MRSHLLNDTAAQYRSRVTEGVRYVAAKLATTDRPFTGYTVDAL
                       SPRIADIDDEPHDTHAAVLEDELDVLRDAVTFHFRYLAHLNCPVTPALLGEAVL
                       SAYNSDLTDWQSGATLIERKIDWTGCAIGLGPAAAGVFTSGGQSNLQALLARE
                       EAKAEADFLIRFASASHFVRKSKALGLGDPADVSI PVDKRMQTVALLARELER
                       CARDGLVPMAVATCGTTFGSDPLPEIAGLCEQGVWVDAAYCCGLLASLKVRD
                       RTIGIRADSVTVVHKSPQPSVSSAVLRDAATLRHATYHAEVLNPRMWQERIPN
                       QVUDKSLQTRERFDALKLWTLRVWAGDGGVLFDEVCDDLAAGCKLLAADPFDVVQ
                       PSLSTLVFRHIPADVDPAEIDRANRYAKALFASGDVAVVAGTKVAGRHYLKFLLNP
                       ETTPADIAVLDLIAGHSAQYIGDSLDRAS"
   misc_feature       1605..2399
                       /gene="SCO2782"
                       /notes="Pfam match to entry PF00282 pyridoxal dec,
                       Pyridoxal-dependent decarboxylase conserved domain, score
                       157.60, E-value 4.6e-45"

```

```

gene
/ gene="SCO2783"
/ note="synonym: SCC105.14"
CDS
/ gene="SCO2783"
/ notes="SCO105.14, probable monooxygenase, len: 413 aa;
similar to SW:RHSB RHIME (EMBL:AF110737) Rhizobium
meliloti rhizobactin siderophore biosynthesis protein RhsE
(EC 1.-.-), 454 aa; fasta scores: opt: 616 z-score:
727.2 E(): 0; 42.3% identity in 430 aa overlap and to
SW:IUCD ECOLI (EMBL:MI8968) Escherichia coli L-lysine
6-monooxygenase (EC 1.13.12.10) IucD, 425 aa; fasta
scores: opt: 802 z-score: 946.5 E(): 0; 31.6% identity in
414 aa overlap"
/ codon_start=1
/ transl_table=11
/ product="putative monooxygenase"
/ protein_id="CA887220.1"
/ db_xref="GI:7544047"
/ db_xref="SPTREMBL:Q9L071"
/ translation="MGIIGIPFNIGLACLTPEVAELNGVFLSKPDPFHWAGMFLDGA
HIQTTPMSDLVTLADPTSPYSFLNLYKEGRLYSFVIRENFYPLRVEYDDYCEWAARK
LSVRPSTTVETVYDERBELYAVATISGTTRARLVLTGTPPHIPDACRGIAGDF
LNSRVRRRAELVKKKSITLVGSGQSAEIYQDLSELDVHGYNWTRSPRRFPL
EYTKLLENTSPDYVYHALPEDTRYLTAEQGLFKGIDGLINEIFDLLYQKRLG
GVPTRLLTNSALTARYADTYTLGROEQGTDFEITEGLVLAATGYRYTEPEFLK
PVRDLRYDSRGNFDIGRNVAVDVTTGGGVFLQAGVHAHSVTSPLDGMGAYRNSCIVR
ELLGREYVPEQSIAPQEFV"
RBS
3881..3885
/ gene="SCO2783"
3894..4448
/ gene="SCO2784"
/ note="synonym: SCC105.15"
3894..4448
/ gene="SCO2784"
/ notes="SCO105.15, probable acetyltransferase, len: 184 aa;
similar to SW:RHSB RHIME (EMBL:AF110737) Rhizobium
meliloti rhizobactin siderophore biosynthesis protein
RhsD, 196 aa; fasta scores: opt: 257 z-score: 326.7 E():
9.2e-11; 33.7% identity in 172 aa overlap and to
TR:AAD46626 (EMBL:AF162771) Pseudomonas aeruginosa
aminoglycoside 6'-N-acetyltransferase Aac, 193 aa; fasta
scores: opt: 175 z-score: 226.3 E(): 3.6e-05; 28.2%
identity in 174 aa overlap"
/ codon_start=1
/ transl_table=11
/ product="putative acetyltransferase"
/ protein_id="CA887221.1"
/ db_xref="GI:7544048"
/ db_xref="SPTREMBL:Q9L070"
/ translation="NSRLSTTTPVGALTLRPVDPLTDVLLHGWLTTPKSAFMMODA
RLVDVERAYMELADEHQAHGLGHDGVPALFTRDYDPAHRELVLGYEPGVDGWHF
LVAPTRPVHGTGTRAVITVWTELEFADPATRRVVVEPDVTNTAVHALNAAVGFVPERE
IQPKKKALLSFTREQFPAKAVSA"
4445..6232
/ gene="SCO2785"
/ note="synonym: SCC105.16"
4445..6232
/ gene="SCO2785"
/ notes="SCO105.16, hypothetical protein, len: 595 aa;
similar to SW:RHSF RHIME (EMBL:AF110737) Rhizobium
meliloti rhizobactin siderophore biosynthesis protein
RhsF, 601 aa; fasta scores: opt: 1094 z-score: 1325.6 E():
0; 32.4% identity in 592 aa overlap and to TR:P94255
(EMBL:U61153) Bordetella bronchiseptica AlnC, 618 aa;
fasta scores: opt: 2037 z-score: 2470.2 E(): 0; 51.5%
identity in 592 aa overlap"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="CA887222.1"
/ db_xref="GI:7544049"
/ db_xref="GOA:Q9L069"

```

```

/db_xref="sp|P091069"
/translation="MSLADAVAHLTTPERWEANRLVVKALAEFTHERLLTTPREPDD
GGGTYVVRSDGDTAYFTATVRALDHQVDAASVTRHRDGAELPLAALDFFELKQ
TLGLSDILPVLYEISSTLSGTCTVLTQKLSARSLARSGDFQAVETGTEGHPCFV
ANNGRLFGFHEVLSYAPSTAGVRLVLAHRSRAAPTAGVGLYESFVRDELGAAT
VDRPHGVLRGDLDPADYLLIPVHPWQWKNLITVFAAEVARGHLVCLGSGDEYLQ
QSIRTFNAPGPHGYKVTALSVNMGFMRLGSAAVMEATPAINDWLRLEIGDPVLK
ETGLSITREAAAGVGRHLEQATDRSPYRKLMAALRESVPFSIREGETLATMASL
VHODHEGASGALIERISGLTPTPEWRLHYLYVPLLSHYFYADVLMYPHGENVILV
LADGVRRVVKDIAEIRVMDPAVLPPVSRIVADVPDDKKLLSTFTDVFDCFRPF
LAANLAEGIVTDEAFWRTVAETREYQESVPELADKFEYDMFAPEFALSCLNRLQ
RDNQVMDLADPSGALQVLGTLKNPLAGR"
/gene

```

gene

```

complement(6302..7909)
/note="synonyms: hexA, SCC105.17c"

```

CDS

```

complement(6302..7909)
/note="SC02786"
/note="SCC105.17c, hexA, beta-N-acetylhexosaminidase
precursor, len: 535 aa; highly similar to TR:085361
(EMBL:AF063001) Streptomyces plicatus
B-N-acetylhexosaminidase, Hex, 561 aa; fasta scores: opt:
2991 z-score: 3318.1 E(): 0: 91.2% identity in 469 aa
overlap and to TR:BA08876 (EMBL:AB015350) Streptomyces
thermophilaceus beta-N-acetylglucosaminidase precursor
NagB, 541 aa; fasta scores: opt: 2856 z-score: 3168.6 E():
0: 76.7% identity in 537 aa overlap. Contains pfam match
to entry PF00728 Glyco_hydro_20, Glycosyl hydrolase family
20 and possible N-terminal region signal peptide sequence"

```

Alignment Scores:

```

Pred. No.: 0.000437 Length: 313800
Score: 654.00 Matches: 637
Percent Similarity: 31.34% Conservative: 264
Best Local Similarity: 22.16% Mismatches: 1051
Query Match: 4.95% Indels: 935
DB: 1 Gaps: 126

```

US-09-522-753-5 (1-2517) x SC0939114 (1-313800)

```

Qy 21 ProHis---SerLeuSerTyProValGluAlaArgThrHisThrAspValGlyLeu 39
Db 113934 CCGACCTTGGCTTGACACACAGCTCCAGCCGATCGTCGGCTGTCGAGGTACAGCAC 113993
Qy 40 LeuGlu---TyrGlnHisHisSerArgAspTyrAlaSerHis---LeuSerProGlySer 57
Db 113994 CTCGGGTCTGTCAGCAGTCGCGCGGATCTCGCGCGCATCGCTGCC-CCAGGAGA 114052
Qy 58 IleleGlnProGlnArgArgProSer-LeuLeuSerGluPheGlnProGlyAenG1 77
Db 114053 GTTCCCGCACCGCACGTCGAGCAGCTCCTGGAGCTGGAGAGTTCGACACAGCGGTCCA 114112
Qy 77 uArgSerGlnGluLeuHisLeuArgProGluSerHisSerTyLeuProGluLeuGlyLy 97
Db 114113 GGTTCCTCGGTTAACGGGGTCCGGGATGCGGTACATGCGGTGTCATCGCGGTAGGAGT 114172
Qy 97 sSerGluMetGluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLe 117
Db 114173 CGATCAGCGCAGGTCCACACAGCGTCTGCGCTGCCGACACACACCGCGGTGCGGT 114232
Qy 117 uLeuArgProSerProLeu-----LeuAl 125
Db 114233 GCGCCAGCGGATCCGCTCCGGGACGGTTCGATGCGCGCACCGCGCGCGCGC 114292
Qy 125 aThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArg----- 139
Db 114293 TCGGCGTCAGGATCGCGGT---GAGCATCTGTGATGTCGTCGATCCCGCGCGCGTTCG 114349
Qy 140 -----SerLeuThrGlyLysLeuGluProValSerProSerProPr 154
Db 114350 GCGGATGTAGCCACCATCTCGCGCGCGCACCGGTGACGACGAGCGCGTCCACCGCCC 114409
Qy 154 oHisThrAspProGluLeuValProProArgLeuSerLys-----GluGluLe 172

```

```

Db 114410 GCACCTG---CCGCGCGCTCGCGCTTTCAGGAAGCCCGCTTCTCTCGCACGCTCGAAGACCT 114466
Qy 172 uileGlnAsnMetAsp-----ArgValAspArgGluIleThrMe 185
Db 114467 TCTCGACGCGTCCAGCTCGATGAGGGTCCGCTCGCTCCACCGCGCATCCGCCC 114526
Qy 185 tValGluGlnGlnIleSerLysLeuLysLysGlnGlnGlnLeuGluGluAaL 205
Db 114527 CCGCGCGGTACCGGTCCCGG---TCCGCGCGGTCCGCGCTCGCGC-TCCTCGCC 114582
Qy 205 aLysProGluProGluLysProValSerPro----- 216
Db 114583 ATGTCGCGCTCTCCATCGCTCCCTCGCTCCCGTACTCCGTAGCAACGCGAGGCCG 114642
Qy 217 -ProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLy 236
Db 114643 CCGCGCAGGCGCGCCCGCCAGGCGACA-----GCAGCGCGCG 114681
Qy 236 sLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLe 256
Db 114682 CCGCGAGCGCGCGGTGAACGCCACCCACCGCGGAGGTCCAGCGGATACGCGCGCCCA 114741
Qy 256 uTyr-----AsnGlnProSerAspThrArgGln---TyrHisGluAsnIleLysI1 272
Db 114742 GCACATAGCTCGCGCGGAGCGGATTCACGAACGCCGCGGAGCGACGACGCTGCCCGC 114801
Qy 272 eAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLy 292
Db 114802 GCACAGGTC-----GAGCGCGAACACGCTCGCGC 114831
Qy 292 sGlnTrpLysGlnLys-----PheCysGlnArgTyrAspGlnLeuMetGluAlaLe 309
Db 114832 GGTACTCGACATCTGCTGCTACCGCGGTAGTGAAGCGGTGGCCACCTCCGAGGCGTCT 114891
Qy 309 uGluLysLysValGluArgIleGluAsnAsnProArgArg-----ArgAlaLysG1 326
Db 114892 GCGCGCGAACTGGAACGC-----CCCGCGCGCACCAACACCGCGCAGAGA 114939
Qy 326 uSerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLe 346
Db 114940 TCCCGCGCGCTGATCAGCGCCACCGGAGCAGCAGCAGCAGCTTCGCGCG---CGTCCAGT 114996
Qy 346 uGlnGluArgMetGlnSerArgValGlyGlnArgGlySerGlyLeuSerMetSerAlaL 366
Db 114997 CGATGTCGACGACACCGCGCGGTACCGGAGCACGACGAGCCCTGACACCGCGCGCA 115056
Qy 366 aArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGluGlnGluAsnLeuG1 386
Db 115057 GCGCGCGCAGCGCAA-----CCGCTCCGCGCACCTGCGCGCAGCACCGCGCGCGAC 115110
Qy 386 uLysGlnMetArg---GlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnG1 405
Db 115111 GCACACGACGCTGTCAGGTCGCGTCCGCGCACCG--- 115147
Qy 405 nArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAs 425
Db 115148 -----GCGGCGCGCGTTCAT---GGAGCGCA 115173
Qy 425 pArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMe 445
Db 115174 TCGCCAGGTC-----GGCCAGCGCGAACA----- 115198
Qy 445 tGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaG1 465
Db 115199 -----CACCCC-----CGACACCGGTACAGGA 115221
Qy 465 uCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgAr 485
Db 115222 A-----CG 115224
Qy 485 gSerTyrArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG1 505
Db 115225 CCACCTCGGCGCAGTAGTCCCGCCAGCGCGTCCACCGCGGAGAACATCAGCAGGATCG 115284

```

```
QY 505 nGlnGlnGlnGlnProMetProArgSerSerGlnGluGluYsAspGluYsGluY 525
D 115285 CCACGAAGTCCAG-----CGCGGTATCGCGAAGTCCCGCAACGCCGTACAGC 115332
QY 525 sGluYsGluAlaGluYsGluGluGluYsProGluValGluAsnAspYsGluAspLe 545
D 115333 CGAACGACGTCCGGTACGC---CATCGTCGACCTGATCCATCCCGCGCATCAGCCGCT 115389
QY 545 uLeuYsGluYsThrAspAspThrSerGlyGluAspAsnAspGluYsGluAlaValAl 565
D 115390 ACGCGCCACCCCTCGACAC----- 115411
QY 565 aSerLySgLYsGlyArgLySgLYsThrAlaAsnSerGlnGlyArgGlySgLYsGlyArg-S 585
D 115412 -----GGCGCCGCGCGTACCCCGCGCGCGCGCGCTCGCGCGCT 115452
QY 585 erMetAlaAsnGluAlaAsnSerGluGluAlaLeThrProGlnGlnSerAlaGluLeuA 605
D 115453 CGCGCGCTCCGCGCGCTCCGCG-----GCCTCTGCCACCTCAG 115491
QY 605 laSerMetGluLeuAsnGluSerSerArg-----TrpThrGluGluMetG 621
D 115492 CCACCTGGACACCACTCGCGTGGCGCGACCTGACACCGCGCGCGCGCGCGCGCGCAGC 115551
QY 621 luThrAlaLySgLYsGlyLeuLeuGluHisGlyArg---AsnTrpSerAlaAlaArgM 640
D 115552 AGCGCCACCCCGCAGGACGCTGGACACAGATACGTCCCGCAGCGATCCCGCTCCCGCAGC 115611
QY 640 etValGlySerLySgLYsThrValSerGlnCysLySgLYsPheAsnTyRlySgLYsArgG 660
D 115612 AGCACGTCCCGCGACCTCGAGCAA-----CGACACCCACCGC 115650
QY 660 lnAsnLeuAspGluLeuGlnGlnHisLySgLYsMetGluYsGluArgAsnAlaA 680
D 115651 AGCGCCGACACCGTCCCGAGACCGCGCGGAACACGTTCAGCGCAGCAGCATCCCG 115710
QY 680 rgArgLySgLYsLYsAlaProAlaAlaAserGluGluAlaAlaPheProValY 700
D 115711 GAGCAGAACAGCCCGCGCACCA-----CGCATCTCGGTCACCCCGGTGCGG 115758
QY 700 alGluAspGluGlu-MetGluAlaSerGly-----ValSerGlyAsnGluGluMet 717
D 115759 TCGAGCAGCCAGAACCGCGCAGCGCACCGCACAGATAGCGCAGCGCAATCCACACCATC 115818
QY 718 ValGluGluAlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCySer 737
D 115819 GCCAACACACCGCCACCGAAGCGCCACACGAGACCTCCCGCGCGCAGCGCCACC 115878
QY 738 GlyProAlaThrValAsnAsnSerSerAspThrGluSerileProSer-ProHisThrG1 757
D 115879 GGGNAGAACAGCAGCACCGAACAGCAGCGCACCAACCGCGCGCGCGCGCGCGCGCGCGCG 115938
QY 757 uAlaAlaLySgLYsThrGlyGlnAsnGlyProLySgLYsProAlaThrLeuAlaAspG1 777
D 115939 GCCCGCC-----CCACGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 115977
QY 777 yProProGly-----ProProThrPr 785
D 115978 TCCGCGGACGGTACAGATCGCGGCGACATCACCGGTACGGATGGCTCCATCAACTCG 116037
QY 785 oProArg-----ArgThrSerArgAlaProIleGluProThrProAlaSerG1 801
D 116038 TCCTCGAACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 116097
QY 801 u-----AlaThrGlyAlaProThr----- 807
D 116098 AACGTACCGCGCTGCGCTGCTATCTCTCGAGCTGCGGCTTCTCGTCCACAGCGGCC 116157
QY 807 ----- 807
D 116158 AGATAGGTGTAAACAGGATCAAAACCGAAGACCGTATTGGTGAACACCCCGCGCGCGGTG 116217
```

```
QY 808 -----ProProProAlaProProSerProSerAl 817
D 116218 GCGCGCCGATACGTCCGTACCGTCCGAATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 116277
QY 817 aProProPro-----ValValProLySgLYsGluGluYsGluGluGluThrAla-AlaA 834
D 116278 CCGAGCCCGCGCATCTCAGCCCTCCCGACCGCGCACCAAGAGCGCAGGACCTAGTG 116337
QY 834 laProProValGluGluGlyGluGlu-----GlnLySgLYsProp 846
D 116338 CGCCCCCGGGGCGCCCTGCCACGGAATTCACGGGTGTCCGTCCGGTGGGAACGGAACC 116397
QY 846 roAlaAlaGluGluLeuAlaValAspThrGlyLySgLYsAlaGluGlu-ProValLySgLYs 865
D 116398 CGCGGTGCGAGAGTCTTCCAGAGAACTCAGAGCTTCCGAAACCGCTACGACGCGAA 116457
QY 866 CysThrGluGluAlaGluGluProAlaLySgLYsAspAlaGluAlaAlaGluAla 885
D 116458 GCA-----GAAGCCGAAACAGAGGTCCGTGCACGAGATGAGCGCAGCGCGCGCA 116511
QY 886 ThrAlaGluGlyAlaLeuLySgLYsAlaGluYs-----LySgLYsGluY 898
D 116512 GCGCCCGAACCGAGGTGGCGCCGAGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 116571
QY 899 GlySerGlyArgAlaThrThrAlaLySgLYsSerSerGlyAlaProGlnAspSerAspSer 918
D 116572 GGAACAGGCGCGCGCGCGGAAACCGCGGAAACCGCGGAAAGAGAGCGCGCGCGCGCGCG 116631
QY 919 -----AlaThrCySerAlaAspGluValAspGluAlaGlu 930
D 116632 CAAGCGCACCGGATGGCGCGCATGTATCCGACCTGCGCATCTGCTGAGCACCGCTCGT 116691
QY 931 -----GlyGlyAspLySgLYsAsnArgLeuLeuSerProArgProSerLeuLeuThrPro 947
D 116692 CGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 116739
QY 948 ThrGlyAspProArgAlaAsnAlaSerProGlnLySgLYsProLeuAspLeuLySgLYs 967
D 116740 CGTGCAGATCCCGCGCGCAACCGCGCTCGCCACCAAGCAGTCCACGCTCTACCTGTACGC 116799
QY 968 GlnArgAlaAlaAlaIleProProIleGlnValThrLySgLYsValHisGluProProArgGlu 987
D 116800 CGACGCGCA-GCCTCTCGCGC-----GCGACGCGCGAGGTCAACCGCGAGAGCG 116846
QY 988 AspAlaAlaProThrLySgLYsProAlaProAlaProPro-----ProProGlnAsnLeu 1005
D 116847 TCGGCTCTCCCGCATCTCCAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 116906
QY 1006 GlnProGluSerAspAlaProGlnGlnPro-----GlySerSer 1018
D 116907 ACTTCTACAGCGAGTCCGCGCATCGACCCCATGGCGATGTCCGGCGCGCGCGCGCGCG 116966
QY 1019 ProArgGlyLySgLYsSerArgSerProAlaProProAlaAspLySgLYsGluAlaPheAlaGlu 1038
D 116967 CCACCGCGAAGCGCAGCAGTCCGCTCCACCATCA----- 117002
QY 1039 AlaGlnLySgLYsProGlyAspProProCyTstPThrSerGly-LeuProPheProValPr 1058
D 117003 -----CCAGCAGTACGTCAAGAACTACTACTCTGCGCGCGCGCGCGCGCGCG 117041
QY 1058 oProArgGluValIleLySgLYsAlaSerProHisAlaProAspProSerAlaPheSerTyRAl 1078
D 117042 AGACCGTCAC-----CCGCAAGGTGAAGGAGTCTTCTCATCTCGATCAAGC 117086
QY 1078 aProProGlyHis-----ProLeuProLeuGlyLeuHisAspThrAlaAr 1093
D 117087 TCGACCGGAGAGAGCAGACGAGATCTTCGAGGCTACCTCAACACAGCTTCTTCG 117146
QY 1093 gProValLeuProArgProProThrIleSerAsnProProLeuIleSerSerAlaLy 1113
D 117147 G-----CGCGCGCG-CCTACGCGCATCCAGCGCGCGCGCGCGCGCGCGCTACTTCGCGGTC 117196
QY 1113 sHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLe 1133
```


Db	117197	GACGCCAAGGACC-----	Db	117992	GTCAGCAGACCGCATCGACTCGGACTCTCCCCCGACACCC-CCGAGCTGACCCCGCTC	118050
Qy	1133	uHisValProTyrSerGluHisAlaLysAlaProValGlyProValThrMetGlyLeuPr	Qy	1441	-----LeuAlaProArgProLeuL	1447
Db	117221	CAGGGCGCTTACCTCGGCTCCTCTCAACGCCCCAGCCAGTACGAGCTCG-----	Db	118051	CCCTCCATCGCCTCGGCTCGCCACCGCCAGCCCTCGACATCGCGGAGGCTACGC	118110
Qy	1153	oLeuProMetAspProLysLysLeuAlaProPheSerGlyValLysGlnGluLeuSe	Qy	1447	ysGluGlySerIleThrGlnGlyThr---ProLeuLysTyrAspThrGlyAlaSerThr	1466
Db	117273	-----TC	Db	118111	GACCTCGCCAAACCGCAGGACACACCTACACGCTGCTCGAGAGGTGCATGAGGA	118170
Qy	1173	rProArgGlyGlnAlaGlyProProGluSerLeuGlyValProThrAlaGlnGluAlaSe	Qy	1466	hrGlySerLysLysHisAspValArgSerLeuIleGlySer---ProGlyArgThrPheP	1485
Db	117275	GCCACCCCGAGAACCGCGCGCGCGAGG-----CCCGCTGGAAGTACGTATG	Db	118171	CGCAAGCAGAGGTGAGCTGCCGCGCGCGCGAGCACACGCGCTCGACCCGCGAGTC	118230
Qy	1193	rValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIlePr	Qy	1485	roProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluAlaCysTyrG	1505
Db	117326	AACGGCATGCGCGACAGGGCTGATCAGCCCTCCAGCGTGC CGCGGGAATTCCCC	Db	118231	CGCGACACACCA-----	118244
Qy	1213	oSerThrArgValPro-----SerAspSerAlaIleThrTyrArgG	Qy	1505	luGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySerIleAlaArgG	1525
Db	117386	AAGCCCAAGCAGTCCACCATCTCCACCAACATGTCCGCGACAGCGCGCTACG-----	Db	118245	-----CGTCCGTCCTCGCAGCGCTCGTCC-----AGAA	118272
Qy	1227	ySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThrIleThrArgIleI	Qy	1525	lyAlaProValIleValProGlu-----LeuGlyLysProArgGlnSerProLeuThr	1543
Db	117438	-----TCGTCAACATCGTCAAGAGTAC	Db	118273	CGCACCGCCACCGCCCGCCGCGCGCTCGTCCCGCCGCGCCGCAAGACCG-----	118325
Qy	1247	eGlyGluAspSerProSerArgLeuAspArgGly-ArgGluAspSerLeuProLysGlyH	Qy	1543	yrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetA	1563
Db	117461	CTGATCGAGAACGATCGTCGACGAGAGAACTCGCGCGCGCGCTACCGCATCACC	Db	118326	-----GCACCCCGAGGAGACACCGCGCTGTTTCCCGGCTACA-----	118367
Qy	1267	isValIleTyrGluGlyLysLysGlyHisValLeuSerTyrGluGlyGlyMetSerValT	Qy	1563	rgGluProThrProArgLeuGlnGluGlySerSerSerSerLysAlaSerGlnAspA	1583
Db	117521	ACCACCTCCAGAGGACAGCAGCAGC-----	Db	118367	-----	118367
Qy	1287	hrGlnCysSerLysGluAspGlyArgSerSerGlyProProHisGluThrAlaAlaP	Qy	1583	rgLysLeuThrSerThrProArgGluLeuAlaLysSerProHisSerThrValProGluH	1603
Db	117549	-----	Db	118368	-----CCCCGA	118374
Qy	1307	roLysArgThrTyrAspMetMetGluValArgValGlyArgAlaIleSerSerAlaSerI	Qy	1603	ishisProHisProIleSerProTyrGluHisLeuLeuArgGlyValserGlyValaAspL	1623
Db	117550	-----CTTCGTCAAGCGCTGACGACAGCTGATGTCCTCCAGCTCGAC	Db	118375	CTTGCGCA-----CGTCTCGTCCCG-----TCAT	118398
Qy	1327	leGluGlyLeuMetGlyArgAlaIleProProGluArgHis-----	Qy	1623	eutyArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProL	1643
Db	117593	AAGAGGAACGCAAGTTCGACACTTACCTCGCGCGCGCGCGCTCGTCCGACCCGAG	Db	118399	GGCGCAGGACCCGAGACCCCGCCACAGCCCTGTACGCGCCATGGCCCTCGCCCG	118458
Qy	1341	-----SerProHisHisLeuLysGluGlnHis-----	Qy	1643	euAspAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProH	1663
Db	117653	ACCGCAAGTCTCGCGGTGTACACGGCATCGACTACGTCAAGCAGTACACGCCCAAC	Db	118459	TATCAACGGCGCGCGCG-----	118475
Qy	1351	-----IleArgGlySerIleThrG	Qy	1663	isLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgG	1683
Db	117713	GCGCTCGCGGTGACTTCCAGGTGCTCCAGCTTCAAGCCCTTCGTGTTTCAGTCCGCG	Db	118476	-----CCCCCGCGAGATCTGGAGCCAGTACACCGCAACCGCTCAAGGGCGAACCGCG	118530
Qy	1357	lnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgGluAlaLysL	Qy	1683	lnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisAsnThrAlaThrAlaM	1703
Db	117773	GTGGAGAACCTCCGAGACGACGAGCGACATGAT-----CAACCCCAACACCCCG	Db	118531	CGCG-----ACTTCGACCTCCAGCTCCAGGACCGCGCGCGCGCTCAC	118572
Qy	1377	euLeuLysArgGluGlyThrProProProProProProProProProProProProPro	Qy	1703	etAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuA	1723
Db	117827	TACGACGGCACCGAGAACCGCCCGTCCAGGGCTGGAGCGCGCGCTACGCGCCCGGAG	Db	118573	CGCGCCCCCGAGCA-----CGCAGCGCTCGAGCAACCGCGCGCGCGCGCGA	118617
Qy	1397	yrLysThrGlnAlaLeuLeuProLeuLysLysLysProAlaHis-----	Qy	1723	snTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeuProValL	1743
Db	117887	AACGA-----GGACCAACCGGACTACGGCCACGTCCCGTCCCGAGCGGCACC	Db	118618	CGAGACGGCGCGCGGAGCGCTCCCGACGAGCGCGGCG-----ACGCCGACGAGGACGGCAC	118674
Qy	1412	-----GluGlyLeuValAlaThrValLysGluAlaG	Qy	1743	euValProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProT	1763
Db	117935	GACAAGTCCATCAACCGCTGTACGCCCCAGATGGCGCTCGACGCTCGGCTCGGACAA	Db	118675	CGGCTCGGAACCCCGCGGACTCCCGACGCGGACGACCCCGCGCGG-----ACGACGG	118731
Qy	1422	lyArgSerIleHisGluIleProArgGluGluLeuArgHisThrProGluLeuPro-----	Qy	1763	hrAlaProGlnProPheSerSerArgHisSerSerSerProLeuSerProGlyGlyProT	1783
Db	117991	-----G	Db	118732	CGGACCGAGTCCGACACCCCGCTCGACCGCGCGGAGACCCCG-----TCCACCGCGCGCGCCCA	118790

Qy	1783	hrHisLeuThrLysProThrThrThrSerSerSerGluArgGluArgAspArgG	1803	Qy	2063	erLeuThrHisAspLysGlyLeuProLys---	HisLeuGluGluLeu-AspLysSerHis	2081
Db	118791	GC-----ACCGAGCCGGCCCCACCGGCGCGCGAC-----	118822	Db	119701	CGAGCAGCGGTCGTGCAGCATCCGATGGCAGCTTGTCTGCTCTGGGAGCTGCTCCT	119760	
Qy	1803	luArgAspArgAspArgGluArgGluLysSerIleLeuThrThrThrThrValGluH	1823	Qy	2082	LeuGluGlyGluLeuArgProLysGln-----	ProGlyProVal	2094
Db	118822	-----	118822	Db	119761	TTCTCTTGGCGCTCAGCCCTGAACCTACTGCCCGCGCTCAGCGCTTCCGCCGCGGCT	119820	
Qy	1823	isAlaProIleTrpArgProGlyThrGluGlnSerSerGlySerSerGlyG	1843	Qy	2095	LysLeuGlyGlyGluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnPro	2114	
Db	118823	-----TCGGCGGGCGAGCACTCTCGAAACGGCGCACCGGGCGCGGGCGG	118868	Db	119821	GCGGCCGAGGCGAGCAGCCCCACGA-CCCCGGCCAGCAGCGCGGCGGACGCC	119879	
Qy	1843	ly-----GlyGlyGlySerSerArgProAlaSerHisSerHisAlaH	1858	Qy	2115	Ser-----SerSerProLeuLeuGlnThrAlaProGly-VaLlysG	2128	
Db	118869	GAGACACCGAGCGCGGGCGGGGACTCGAACACC-----GGAT	118910	Db	119880	GCTCCATCCGCGGGCGCGCTCTGCTCA-----CGAACTCGCGCTTCACTCGCTGA	119933	
Qy	1858	isGlnHisSerProIleSerProArgThrGln---AspAlaLeuGlnArgProSerV	1877	Qy	2128	yHis-----GlnArgValValThrLeuAlaGlnHisI	2139	
Db	118911	TCCCCGGCATCCCGTCAACCCAGCGGTCAAGTACCGGAGTCCGCTTGAGACCCACCA	118970	Db	119934	CCACCCGATTGACCTGGACATACGCCCGCCGAGCGTGTGATCGACACTGGAGGCGACCT	119993	
Qy	1877	allLeuHisAnthr---GlyMetLysGlyIleIleThrAlaValGluProSerLysProT	1896	Qy	2139	eserGluValIleThrGlnAspTyrThrArgHisPro-----	2152	
Db	118971	CGCGCACACGACGACAGATGAAGAAGATCCGGCGCGGTTCGCGGCTCCCCCAGCA	119030	Db	119994	TGGCCTTCGCGTCGTCGACGATCGTTCGGGTGCACCCGCATCCCGAGCTCGTCGAGCG	120053	
Qy	1896	hrValLeuArgSerThrSerThrSerProValArgProAlaAlaThrPheProProA	1916	Qy	2153	-----GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSe	2170	
Db	119031	CCACCATGCCAGACCGCGCC-----CGGCGCGACCGATGCC-----	119071	Db	120054	TCTCGGCCAGCACCTCGCGCGCGGTTCGA-TGTCGCGCTCGATCTGCGCGCGGCTCTG	120112	
Qy	1916	laThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrPro-----	1930	Qy	2170	rCysProVal-LeuAspLeuArgProPro---SerAspLeuTyrLeu-	2185	
Db	119072	--ACCACACGCGGTA-GGCGGTACC---GATGGCGAGCGTCCGCGCGGTACGACAGC	119125	Db	120113	ATGTCGAGCGTGTCCGCCACCGTACGCCCTCCGAAGTCCGCTCTGACGCTGTCCCGGACAG	120172	
Qy	1931	-----ThrLeuMetGluProValLeuLeuProLysGluA	1942	Qy	2186	-----ProPro-----ProAspH	2190	
Db	119126	AGCATCATGCTCCGACGATCCCGGCACCCGTCGAACACACTGGGGCCACAGCCGGTGAA	119185	Db	120173	TCTGTGAGTACTTGTCTTCCGCACCTGTCAGGACCCCGGTACGCTGGATCGATGAGC	120232	
Qy	1942	laProArgValAlaArgProGluArgPro-----	1951	Qy	2190	isGlyAlaProAlaArgGlySerProHisSerGluGlyLysArgSerProGluProA	2210	
Db	119186	CCGTCGGTGTACTTCTCCGATCGACCGACCGACCTCGAGCAGACCGGCGGACGAACAC	119245	Db	120233	GAGCGACTCCAGCCGGGACCC-----GGCCCCCGCTTACCCCTCCCG	120277	
Qy	1951	-----	1951	Qy	2210	snLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProProGluG	2230	
Db	119246	AGAACCCACGCCATGACGGGCACCTCCGTCGAACGAACCTGAACGGTGGTGGTCTTTG	119305	Db	120278	GAGCGCGACGGCAACGAGGT-----GTCCCTGTCCGACACCA	120315	
Qy	1952	-----ArgAlaAspT	1955	Qy	2230	lyMetThrGluProGlyHisSer-----ArgSerAlaValTyrProLeuLeuTyrA	2247	
Db	119306	CCTTCAGCCCGTACGGCGCGTCTGTCGGGGTGTCTTCCCAAGGTAGCAAGGACGGGA	119365	Db	120316	-----GGGCGCGAAGTCTATCTACTTCTACCTCGCGCGCTTACCCCGGCTGCACG	120370	
Qy	1955	hrGlyHisAlaPheLeuAlaLysProProAlaArgSerGlyLeuGluPro---AlaSerS	1974	Qy	2247	rgAspGlyGluGlnThrGluProSerArgMetGlySer---LysSerProGlyAsnThrS	2266	
Db	119366	AAGGGGTGCTGACCTTGATCACCAGCCCTGAAGTCCGACCTGGACCCGCTGATTCA	119425	Db	120371	AAGCAGGCCT-GCGACTTCACGACACACCTGGAGCTGCTGACCGCGCGGGTACGACGT	120429	
Qy	1974	erProSerLysGlySerGluProArgProLeuValProProValSerGlyHisAlaThrI	1994	Qy	2266	er-----GlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVal	2284	
Db	119426	ATACAGCCCGGTGAGTCTCCGACCCCTCGAACCCGCTCGCGCCACCGCGTGCAGATC	119485	Db	120430	CATCGGCATCTCCCCCGACA-----AGCCGGAGAAGCTGGCCCAAGTTCCGCG-	120476	
Qy	1994	leAlaArgThrProAlaLysAsnLeuAlaProHisAla-----	2007	Qy	2284	ysSerLysLysGlnGlnIleAsnLysLeuAsnThrHisAsnArgAsnGluProGluT	2304	
Db	119486	ACGCTCGCGATGAGCACGT-----AGCCACACACCCCGCACCTCGACCTCGGACGGTC	119539	Db	120477	-----AGGCGGAGTCCCTGAA	120492	
Qy	2008	-----SerProAspProProAlaProProAlaSerAlaSerAspProHisA	2023	Qy	2304	yrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyL	2324	
Db	119540	CTCCGGTCTGTACAGCACCCGCGTCCCGGCTCCAGGTCGATGCTTCTGAC-----	119592	Db	120493	GGTCAACCTCTCGCG-----ACCCGACAGAA	120522	
Qy	2023	rgGluLysThrGlnSerLysProPheSerIleGlnGluLeuLeuArgSerLeuGlyT	2043	Qy	2324	euMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluA	2344	
Db	119593	-----CGACCGCACCCACC-TCGGCCCGGACCGCGCTTCGCGACCGCGCGCG	119640	Db	120523	GGTCTCGAGCGGTACGGCGCTTCGGCGAGAGAACTAG-----	120566	
Qy	2043	yrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSerSerProS	2063	Qy	2344	laIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSerProProLeuS	2364	
Db	119641	TCGGGGGATCAGGATCCCGCCCGGAAACCGCGCTCGCCCTCACCGGTCTCTGCGGCA	119700	Db	120567	-----GCAAGACCTACATGGGCGCTCATCCGCTCCACGAT	120600	
				Qy	2364	er-----AlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProA	2379	

```
Db 120601 CGTGTGACAAACAGGCAAGTCGAACGGCCCTGTACACGTCGCGCCA---CCGG 120657
Qy 2379 laAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyG 2399
Db 120658 CCAGTGGCCA-----AGATCATCAAGGACTTGGGGATCTCAGCGCG 120699
Qy 2399 ly-----GlyGlyLysAlaLysValSerGlyArgPro----- 2409
Db 120700 CTGCCCCAGCAGTAGCATGCGCGGCAGCAGCGCC-----GGCGCGCGTGTGGA 120750
Qy 2410 -----SerSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProp 2428
Db 120751 ATTGGCAGTCACGCTGGGTTTAGGTCCAGTGGGGTAACTCCCGTGAGGGTT-CGAGTCC 120809
Qy 2428 roSerValSerSerValHisSerGluGlyAspCysAsnArgAlaGlyThrProLeuThrAsnA 2448
Db 120810 CTCGGCGGCACCGTA-----CGTTTCCGATGCTCAGCC 120845
Qy 2448 rgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyAsnProLeuI 2468
Db 120846 GAGCAGCTCGCGCACCAACCGGCACCGCCCGGACGCT----- 120886
Qy 2468 leMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAla---- 2486
Db 120887 -----TGCGCGGCTGGGTGATCGCGTTCTTCTCTCGCGCGTCTCAGTTCCGCGCACGT 120938
Qy 2487 --GlySerGlyProLeuAlaGly 2493
Db 120939 TCGGCTCTCGCCCTCCGGCTGGA 120961
```

Search completed: April 16, 2004, 10:50:23
Job time : 42585 secs

This Page Blank (uspto)